Virtue Consequentialism

BEN BRADLEY
Syracuse University

Virtue consequentialism has been held by many prominent philosophers, but has never been properly formulated. I criticize Julia Driver's formulation of virtue consequentialism and offer an alternative. I maintain that according to the best version of virtue consequentialism, attributions of virtue are really disguised comparisons between two character traits, and the consequences of a trait in non-actual circumstances may affect its actual status as a virtue or vice. Such a view best enables the consequentialist to account for moral luck, unexemplified virtues, and virtues and vices involving the prevention of goodness and badness.

I. INTRODUCTION

In her famous paper 'Virtues and Vices', Philippa Foot writes that 'virtues are in general beneficial characteristics, and indeed ones that a human being needs to have, for his own sake and that of his fellows'.[1] Foot here expresses a view about virtue that we might call 'virtue consequentialism'. It has been held, in one form or another, by Hume, Bentham, and G. E. Moore, and more recently by Linda Zagzebski, Judith Thomson and Julia Driver.[2] It has also been widely criticized.[3] Unfortunately, though the basic idea behind the view seems easy enough to grasp, virtue consequentialism has never been properly formulated. In what follows I try to remedy this situation. I take as a starting point the view presented by Driver in *Uneasy Virtue*, the most detailed and thoroughly defended recent formulation of virtue consequentialism. In light of the problems Driver's view faces, I suggest that the best version of virtue consequentialism is a contrastivist view – that is, a view according to which attributions of virtue are really disguised comparisons between two character traits – and a counterfactualist view – that is, a view according to which the consequences of a trait in non-actual circumstances may affect its actual status as a virtue or vice.

[1] Foot, 'Virtues and Vices', *Virtues and Vices and Other Essays in Moral Philosophy* (Los Angeles, 1978), p. 3.

[2] Jeremy Bentham, *Introduction to the Principles of Morals and Legislation* (New York, 1948), p. 131; David Hume, *An Enquiry Concerning the Principles of Morals* (Indianapolis, 1983), p. 20; G. E. Moore, *Principia Ethica* (Cambridge, 1903), sect. 103; Linda Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge, 1996); Judith Thomson, 'The Right and the Good', *The Journal of Philosophy* 94 (1997); Julia Driver, *Uneasy Virtue* (Cambridge, 2001).

[3] For a recent example, see Rae Langton, 'Virtues of Resentment', *Utilitas* 13 (2001).

© 2005 Cambridge University Press
Utilitas Vol. 17, No. 3, November 2005
doi:10.1017/S0953820805001652 Printed in the United Kingdom

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

Two points before we begin. (1) The goal here is mainly clarificatory. I want to demonstrate the structure of the most plausible and coherent version of virtue consequentialism. I present no arguments against the view that some non-consequentialist element ('having one's heart in the right place', for example) must be part of any correct theory of virtue. However, even those who are not virtue consequentialists might find something of use here. For instance, some think that good consequences are relevant to determining whether a character trait is a virtue, while maintaining that other factors must also be part of the full story.[4] (2) Zagzebski describes a 'pure virtue theorist' who accepts that (necessarily?) virtues have good consequences, but denies the apparently stronger claim that it is *because* it has good consequences that a trait is a virtue.[5] There is no good reason for anyone ever to hold such a combination of views. Nevertheless, in what follows I am concerned only with properly formulating the weaker view, apparently shared by some 'pure' virtue theorists, that there is a necessary connection between being a virtue and being a trait with good consequences.

II. DRIVER'S VIEW

Driver states her view in the following way: 'Specifically, the account that I want to propose is an objective consequentialist account of the virtues, which would define moral virtues as character traits that systematically produce more actual good than not.'[6] Let us begin with this formulation of virtue consequentialism:

VC1: V is a virtue (in the actual world) iff V is a character trait that systematically produces more good than not in the actual world.[7]

Some details of Driver's view need to be explained. First, note the term 'systematically'. It is not enough, on Driver's view, to produce

[4] Zagzebski holds that it is a necessary condition for a trait to count as a virtue that it is beneficial to others (*Virtues of the Mind*, p. 101); Foot and von Wright hold that in addition to having good consequences, virtues are 'corrective' (Foot, 'Virtues and Vices', p. 8; von Wright, *The Varieties of Goodness* (London, 1963), ch. VII). Brad Hooker maintains that while it is controversial whether virtues are intrinsically good, just about everyone believes virtues are at least *instrumentally* good ('The Collapse of Virtue Ethics', *Utilitas* 14 (2002), p. 22), while Zagzebski says it is part of the 'pretheoretical notion' of virtue that virtues have good consequences (Zagzebski, *Virtues of the Mind*, p. 99). Hooker and Zagzebski might be overstating the case, but if they are right, then what I show here is that there is at least one belief almost everyone has about virtue that stands in need of serious clarification.

[5] Zagzebski, *Virtues of the Mind*, p. 99.

[6] Driver, *Uneasy Virtue*, p. 68.

[7] I presume Driver thinks that whether a trait counts as a virtue in some other world depends on its consequences in that world, not in the actual world.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

good consequences on one or two occasions in order for a trait to count as a virtue. If that were enough, every character trait would be a virtue, since even traits we all acknowledge to be vices occasionally and luckily result in good consequences. Thus, the systematicity requirement minimizes the effects of moral luck.[8] Second, note the phrase 'more good than not'. This is Driver's way of formulating her view in a non-maximizing way. A character trait need not have the best possible consequences in order to count as a virtue; it merely needs to produce 'more good than not'.[9] Third, note the qualification 'in the actual world'. The point here is to distinguish her view from views according to which a trait's actual status as a virtue depends on its consequences in counterfactual circumstances. Finally, Driver remains neutral as to the question of what things are intrinsically good. I follow Driver in this respect; the reader may plug in his or her favourite substantive axiological view.

Driver repeatedly says that in order to count as a virtue, a trait must produce more good than not. But what exactly does this mean? To what does 'not' refer? The clue lies in the following passage: 'On this view, with respect to character traits, one is comparing the good produced with the bad produced and judging the good to be greater.'[10] This suggests that we might best state Driver's view as follows:

> VC2: V is a virtue iff V is a character trait that produces more good in the actual world than it produces evil in the actual world.

VC2 is not an acceptable version of virtue consequentialism. Consider a person named 'Downer' who lives in a world full of happy people. Downer engages in various activities designed to lessen their enjoyments; he replaces the pleasures they would have otherwise had with less pleasant pleasures. For example, he invents new, less tasty foods and hoodwinks people into thinking those foods are healthier than the tasty foods they like better. He never causes them any pain – he does not have the stomach actually to hurt anyone – but he does cause them to be less happy than they would have been otherwise. There are no intrinsically bad consequences to appeal to here. Downer brings a lot of intrinsic good into the world, consisting of all the moderately pleasant experiences people have as a result of his actions.[11] He brings no intrinsic evil into the world. Downer's malice produces more good

[8] Though not enough; see section V.

[9] This is in contrast to the maximizing consequentialist view proposed by Moore in Section 103 of *Principia Ethica*.

[10] Driver, *Uneasy Virtue*, p. 122n.

[11] For illustrative purposes I assume that at least some pleasures are intrinsically good. It makes no difference to the argument.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

than evil in the actual world.[12] Thus, VC2 entails that Downer's malice is a virtue. Downer's malice is obviously a vice; yet according to Driver, 'vices . . . produce bad states of affairs'.[13] Downer's malice produces no bad states of affairs (it merely prevents good ones), so it does not count as a vice on Driver's view.[14]

Driver could argue that a character trait such as Downer's must have some intrinsically evil consequences. But I see no reason to think this must be so. Even if it were, in order for the character trait not to count as a virtue, its intrinsically evil consequences would have to be equal to or greater than its intrinsically good consequences; it is easy enough to imagine this not being the case.

The case of Downer is a problem for VC2, but it sheds light on an important consequentialist idea: preventing goodness is just as bad as producing evil, and preventing evil is just as good as producing goodness.[15] Downer's character trait is reprehensible on consequentialist grounds, not because of the states of affairs it brings about in the actual world, but because of the states of affairs it prevents. If we are concerned with evaluating the consequences of Downer's malice, surely we must compare the results of the malice with what the malice prevents – namely, all the more pleasant experiences people would have been having otherwise.

III. CONTRASTIVISM

In light of the example of Downer, a better consequentialist approach to evaluating character traits is to say that a virtue is a character trait that makes the world better than it would have been had that trait not been instantiated. If we were to incorporate this approach without making other changes to Driver's view, we might get the following view:

VC3: V is a virtue iff V is a character trait that systematically produces a greater balance of intrinsic good over intrinsic evil in the actual world than the absence of V would systematically produce.

VC3 apparently accounts for the case of Downer, since his malice makes the world worse than it would be if he were not so malicious.

[12] If traits are universals, Downer's malice is identical to everyone else's malice, i.e. it is malice itself. Malice itself could have consequences beyond those produced by Downer's particular instantiation of it. So to ensure that malice brings more good than evil into the world, I stipulate that Downer is the only malicious person.

[13] Driver, *Uneasy Virtue*, p. 74.

[14] Of course, Downer produces less good than he could have; but that is not enough for him to count as malicious on Driver's view, since Driver is not a maximizer.

[15] See Earl Conee, 'Instrumental Value without Intrinsic Value?', *Philosophia* 11 (1982), and Ben Bradley, 'Extrinsic Value', *Philosophical Studies* 91 (1998), for further discussion of the value of prevention.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

But there remains an important difficulty. A world where V is absent must be a world in which people possess some other character trait instead; but which one?

Since there are many ways for the inhabitants of a world to lack a given character trait, there may be more than one answer to the question. For example, when considering whether being fairly honest has good consequences, we may be comparing a world where people are fairly honest to a world where people are very honest, or totally dishonest, or mostly dishonest, or dishonest when it suits them. Whether a character trait turns out to have good consequences or not may depend on which comparison we make. The world might be better if everyone were fairly honest than it would be if people were dishonest whenever it suited them to lie, but it might not be better than if people were very honest.

Every non-maximizing consequentialist theory faces this difficulty. It seems like a good idea to abandon maximizing because maximizing accounts of virtue are highly demanding; it seems odd to refuse to call someone virtuous simply because she does not have the best possible character trait she could have. For example, Driver describes a person who resists a great deal of Nazi torture without giving in, but succumbs 'when driven to the edge of madness by the threat of being eaten by rats'.[16] We apparently must give up maximizing to call such a person virtuous.[17] But what do we put in the place of maximizing?[18]

Those who wish to hold a non-maximizing consequentialist theory of virtue should say that attributions of virtue are always contrastive in nature; linguistic context picks out an appropriate comparison trait. The trait of being fairly honest has good consequences, and so counts as a virtue, when compared to the trait of being mostly dishonest (a comparison made salient by asking questions such as 'is it good to be fairly *honest*?'); it has bad consequences, and counts as a vice, when compared to the trait of being mostly honest (a comparison made salient by asking questions such as 'is it good to be only *fairly* honest?'). No character trait has the property of being a virtue full stop, because there is no such property. Instead of attributing such a property to a trait, we should say that it is a virtue to have one character trait C1 rather than

[16] Driver, *Uneasy Virtue*, p. 73.

[17] It is not clear that this example really works. It seems plausible to suppose that in the rat-eating example, it would be better if the person caved in to the Nazis immediately. It seems like a case where maximal courage has less than maximal consequences. We must appeal to the consequences of courage everywhere to see it as a virtue. Thanks to Eric Moore for discussion of this case.

[18] The difficulties here closely parallel those pointed out for consequentialist theories of good action by Alastair Norcross in 'Good and Bad Actions', *The Philosophical Review* 106 (1997).

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

some other trait C2. When we call a character trait a virtue, context must provide another character trait to fill the second place of the relation (C2). This trait must be drawn from a contrast class of traits the possession of which is incompatible with the possession of C1. Call this view 'virtue contrastivism'.[19]

Other views are possible. For example, we might think that for each class of incompatible character traits, there is a mediocre character trait that always serves as the benchmark for virtue; those members of the class with better consequences than the benchmark are virtues, those with worse consequences are vices. But this option requires us to locate a benchmark, and any choice of benchmark will inevitably be arbitrary.

A more promising alternative to contrastivism is a view we might call the 'close enough' view. We might think of the difference between the close enough view and contrastivism in this way: while contrastivism makes virtue like tallness, on the close enough view we understand virtue in something like the way we understand flatness. We accept attributions of flatness to things that, strictly speaking, are not truly flat. But they are close enough. Similarly, we count many people as virtuous who really are not virtuous, but are close enough. The person in Driver's rat-torture example is close enough to being perfectly virtuous to make attributions of virtue to him assertible (though strictly speaking false). There would certainly be some vagueness concerning what counts as being close enough, but that is no objection, since vagueness is everywhere.

I find contrastivism more appealing than the close enough view for two reasons. First, it just seems wrong in the rat-torture case to say that the person is strictly speaking not virtuous, in a way that it does not seem wrong to deny that the table is strictly speaking flat. The reason attributions of virtue to such a person are assertible is that they are true. Second, the close enough view is saddled with an assumption that the contrastivist view is not; namely, that there is a maximally good character trait for each class of alternative traits.[20] It seems possible that for every character trait, there is a better trait someone might have had instead. This is compatible with contrastivism but not the close enough view. In the absence of a reason for thinking

[19] Virtue contrastivism is relevantly similar to epistemic contrastivism, as developed by Jonathan Schaffer, 'From Contextualism to Contrastivism', *Philosophical Studies* (forthcoming). See Schaffer's article for a detailed explanation of how contrastivism allegedly differs from contextualism. For my purposes here, it does not matter whether there really is an important difference between contrastivism and contextualism. If there is not, then the view presented here is just a plain old contextualist view.

[20] Aquinas makes this assumption in the 'fourth way' (*Summa Theologica*, Part I, Q.2, a3); see Rosalind Hursthouse, *On Virtue Ethics* (Oxford, 1999), p. 141, for a prominent recent example.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

that every class of alternative traits has one that is maximally good, it seems contrastivism has an advantage over the close enough view. Perhaps there is some other argument for the close enough view over contrastivism, but for now I tentatively conclude that virtue consequentialists should be contrastivists.

Even on the contrastivist view, there are traits that might be called virtues unequivocally. Consider the character trait of being at least a little bit honest.[21] The only way to lack this character trait is not to be honest at all; so only one comparison is possible. This might give some small comfort to those who find the contrastivist view too flexible.

It might be thought that virtue contrastivism is so counterintuitive that if virtue consequentialists are committed to contrastivism, there is reason enough to reject virtue consequentialism. But it is not just consequentialists who have reason to formulate their virtue theory in the contrastivist way. Some of the same desiderata that lead the consequentialist to be a contrastivist might also lead, for example, the Aristotelian to be a contrastivist: namely, avoiding maximizing and avoiding arbitrary thresholds. Given a range of alternative character traits, the Aristotelian must either say that just one is strictly speaking a virtue (the one that is precisely in 'the mean'), resulting in a theory that seems too demanding, or that all those traits that meet some threshold are virtues (those which are sufficiently close to the mean), resulting in a theory that sets an arbitrary boundary. Faced with such a choice, the Aristotelian might find it better to state the view in a contrastivist way: it is a virtue to have C1 rather than C2 just in case C1 is closer to the mean than C2. Virtue contrastivism is neutral between substantive theories of virtue.

IV. INDIVIDUALISM OR UNIVERSALISM?

We have been talking about the consequences of the possession of a character trait, but by whom? By an individual, all of humanity, or some subset of humanity? Call individualism the view that what is relevant to deciding whether it is a virtue for S to have trait T1 rather than trait T2 is what would happen if S were to have T1 rather than T2; call universalism the view that what is relevant to deciding whether it is a virtue for S to have T1 rather than T2 is what would happen if everyone were to have T1 rather than T2. The difference between individualism and universalism is not trivial. There could be a character trait that would be good for one person to have, but would not be good for everyone to have. Conversely, there could be a trait

[21] Thanks to Andre Gallois for the example.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

that would be good for everyone to have but not for one person alone to have. To illustrate both sorts of traits, consider Costanza, whose efforts to help or harm others routinely backfire. The backfiring is not fluky; it is a lawlike backfiring, the result of his bumbling disposition. A world in which everyone is benevolent would likely be a better world than this one; but a world in which Costanza is benevolent (and everyone else is just as they actually are) would probably be a worse one. A world in which everyone is malicious would be a worse world than this one, but a world in which Costanza is malicious (and everyone else is just as they actually are) would likely be a better one. Would Costanza be more virtuous if he were malicious or benevolent?

In Costanza's case, universalism seems more appropriate; it seems wrong to say that it is a virtue for him to be malicious rather than benevolent, even though it would have good consequences in his particular case. However, it is not hard to think of cases that favour individualism. If I want to know what character traits would be best for me to have, it seems relevant to consider my circumstances; one circumstance that is often relevant is what traits other people actually have. For example, it seems wrong on consequentialist grounds to attribute virtue to a diplomat who produces bad results by being exceedingly honest rather than strategically deceptive, even if things would be better if everyone were as honest as he is.

Perhaps it is best to try to account for what seems right about both individualism and universalism. One way to do this is to make attributions of virtue relative to a person or population; it is not a virtue to have V1 rather than V2 full stop, but for a person or for some people. Thus it could be a virtue for Costanza to have V1 rather than V2, but a vice for humanity as a whole to have V1 rather than V2. Of course, this strategy of accommodation is somewhat unsatisfying; if we want to know whether Costanza's traits are virtues, we get different answers depending on whether we are looking at the consequences of the traits when possessed only by him or by everyone. So we do not get an unequivocal answer to the question of whether Costanza is virtuous.[22]

Nevertheless, the accommodation strategy has advantages; for one thing, it seems to give us a way to account for the notorious case of the courageous villain.[23] The courageous villain seems problematic for virtue consequentialism because a villain often brings about worse results by being courageous than by being cowardly, suggesting that it is a virtue for him to be cowardly rather than courageous. But it is always a virtue to be courageous rather than cowardly (or so the objection goes).

[22] Thanks to Erik Wielenberg for discussion of this point.
[23] See Foot, 'Virtues and Vices', pp. 14–18, for one discussion of this sort of case.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

While we do think of courage in the abstract as a virtue, in the case of the villain, it is not so clear that this is the correct thing to say. His courage makes him even more villainous. By accommodating universalism and individualism, VC tracks the ambivalence we feel about the courageous villain. It is probably a virtue for humanity as a whole to be courageous rather than cowardly, but not for the villain in particular. This is one of those cases in which it is correct both to assert and to deny (though not with respect to the same group of people) that courage is a virtue. Thus I see the courageous villain as providing some support for the accommodation strategy.

V. COUNTERFACTUALISM

The second important difficulty for Driver's theory is her commitment to what she calls 'actualism'. According to Driver, 'actualism holds that the moral quality of actions and character is determined by actual consequences, or consequences in actual circumstances'.[24] We have already seen, in the example of Downer, that actualism so stated must be false. A consequentialist must look at the values of non-actual circumstances to determine whether a trait is a virtue. But there is still a sort of view we might call 'actualism' that is worth considering. According to this view, to determine whether a trait is a virtue, we compare its consequences in the actual world with the consequences that would have resulted from the possession of an alternative trait. Actualism so formulated may be contrasted with counterfactualism, which Driver characterizes as follows: 'a moral virtue is a character trait that would systematically produce actual good under normal circumstances (i.e. in possible worlds very close to our own)'.[25] Driver's arguments for actualism are based on its contrast with this version of counterfactualism. This is problematic, since a different version of counterfactualism is more plausible. In any case, Driver's arguments fail to identify a genuine problem for counterfactualism even when it is formulated in the way she suggests.

First, some clarification. Ordinarily, when we say that something would have happened, we mean it would have happened if something else had (or had not) happened. In light of Driver's remarks about the view, I interpret her formulation of counterfactualism as the view that a virtue is a character trait that would have good consequences if circumstances were only slightly different from the way they actually are (even if, by sheer luck, it happens to have bad consequences in the actual world). (Foreshadowing: my preferred

[24] Driver, *Uneasy Virtue*, p. 77.
[25] Ibid., p. 78.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

version of counterfactualism is the view that a virtue is a character trait that would have good consequences if people were to possess and exercise it. More on this in section VI.)

Driver's main complaint about counterfactualism is found in the following passage:

> Suppose, for example, that Sally would have had bad traits if she had not been raised by her mother, who, it turns out, did raise her only through amazing luck – the mother was almost run over by a truck but avoided death through an amazing fluke. Well, in worlds very close to this one, Sally is a bad person. Her high spirits become disruptive. Does it make sense to say that her high spirits are a vice in this world because in worlds close to ours, though not in ours, they are disruptive? One can also reformulate the case to make it less particular to Sally. Suppose that high spirits produce good in the actual world through some cosmic fluke, yet fail to in nearby possible worlds. Sally's high spirits are still virtuous.[26]

Driver seems to be suggesting that the counterfactualist must call Sally's 'high spirits' a vice, since in 'worlds close to ours', that trait is disruptive. One complication that clouds the issue here is the fact that the example is intended to involve the consequences of a particular character trait: high spirits. But presumably it is not merely her high spirits that have good or bad consequences; her high spirits are combined with other traits to produce good consequences in the actual world, and are combined with still other traits to produce bad consequences in the nearby worlds. In nearby worlds where she has all the traits she actually has, I presume they would most likely have good consequences as well. So the counterfactualist can account for Driver's counterexample by denying that it is a virtue merely to have high spirits without also having other relevant traits.

But the real problem with Driver's argument is the assumption that there is only one correct answer to the question 'which are the nearest possible worlds to the actual world?' and that the answer to that question is 'worlds where Sally's mother was hit by a truck'. But there are other answers, including 'worlds where Al Gore was elected president in 2000', and 'worlds where the Yankees won the World Series in 2003'. It is by now a truism that any one of these answers, or thousands of others, may be right, depending on which respects of similarity are deemed to be important. In worlds where Gore was elected or the Yankees won, Sally's high spirits would still have good consequences, since the other aspects of her character would remain unchanged, barring some unknown connection between virtue and elections or between virtue and baseball. There is simply no good

[26] Ibid., p. 81.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

reason to think that the counterfactualist is committed to the bald claim that Sally's high spirits have bad consequences in nearby worlds.[27]

Furthermore, there are at least three other arguments against actualism. I will call them the Moral Luck Argument, the Unexercised Virtue Argument, and the Uninstantiated Virtue Argument.

A. The moral luck argument

Lucky is a violent psychopath who repeatedly attempts to kill people. Each time he does, something highly unusual happens (e.g. a bird flies in the path of the bullet; a severe gust of wind knocks him over) and his attempt fails. He is not incompetent, just continually unlucky. His murderous tendencies fail to bring about anything bad in the actual world. But if the bird had not flown just there, or if the wind had not blown just that way, he would have successfully killed many people. The actualist apparently lacks the resources to say that Lucky is vicious.

Driver might think that this example is misleading. It is not just the effects of the instantiation of malice in Lucky that are relevant to whether Lucky's malice counts as a virtue; it is the effects of malice throughout the actual world when possessed by anyone ('systematically', as Driver says). Many of those other instances of malice have bad consequences in the actual world, making Lucky's malice (i.e. malice itself, if traits are universals) a vice.

This can help Driver account for some cases of moral luck, but it is inadequate as a general response to the moral luck problem, for two reasons. First, we could stipulate that Lucky is the only malicious person in the world in question, so that the effects of his malice are the only relevant effects. Second, we could also imagine an even more fluky world in which not only do Lucky's attempts at violence fail, but so do everyone else's, owing to convenient birds, gusts of wind, avalanches, or whatever. If the point is to eliminate from consideration traits that produce good consequences merely by accident, we must reject actualism.

B. The unexercised virtue argument

Helper is a generous person. If anyone ever were to need her assistance, she would provide it even at great expense to herself. As it happens, she never encounters a single person who needs her help. (And not because she intentionally tries to avoid getting herself into such situations.) Thus her generosity never gets exercised. Since it never gets exercised, it never has any actual consequences; *a fortiori*, it never has any good consequences. The only way to find some good consequences of her

[27] See David Lewis, *On the Plurality of Worlds* (Oxford, 1986) for more on the idea of 'closeness' of possible worlds.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

character trait is to look at counterfactual circumstances. Actualism entails that her generosity is not a virtue.

Interestingly, Driver discusses cases like this and almost recognizes that they constitute a good reason to reject actualism:

> In a very repressive society, it may be the case that many people possess virtuous dispositions yet lack the opportunity to display the requisite behavior. For example, loyalty would have little scope in a society that discouraged friendship. Although actual good consequences may not be produced by these dispositions in a given instance, these people still have the virtues because the dispositions would produce the requisite behavior in the appropriate context.[28]

Driver seems to admit that we need to look at more than just actual consequences when evaluating a character trait. Does this contradict her officially stated position? She does not seem to think so:

> Counterfactual claims are quite relevant to virtue evaluation. If good would be produced by the agent but for some fluke, or bad luck in the actual world, the agent still has moral virtue as long as the relevant trait produces good systematically in the actual world.[29]

Driver seems to be saying that counterfactuals are not relevant at all; what are relevant are the consequences of the trait throughout the actual world. So in the case of Helper, Driver would say it is not just the consequences of Helper's exercise of generosity we need to look at, but the consequences of generosity everywhere in the actual world; when we do so, we find many good consequences. But as before, this reply is inadequate. There is no reason to stipulate that others have Helper's character trait, or that if they did, their exercise of that trait must have good consequences.

C. The uninstantiated virtue argument

Diogenes the Cynic is purported to have said 'I am looking for an honest man'. If Diogenes was looking for a person with some virtue, and he failed to find one because no such person existed, it would seem that according to actualism, the character trait he was looking for was not a virtue after all – though it might have been a virtue had he only found it. Consider a level of benevolence exceeding that of Mother Teresa; call this trait 'super-benevolence'. Since super-benevolence is uninstantiated, it has no consequences in the actual world; *a fortiori* it has no good consequences. Thus the actualist must deny that super-benevolence is a virtue. It is an odd feature of actualism that when determining what the virtues are, we first have to look at what

[28] Driver, *Uneasy Virtue*, p. 75.
[29] Ibid., p. 83.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

character traits people actually happen to have, since those are the only ones with actual consequences.

The actualist could respond by noting that super-benevolence would be a virtue, if only it were realized. But this suggests that by being the only person to instantiate this trait, an agent would be making it a virtue, since it would not have been a virtue if he had not had it. He would be creating a new virtue rather than merely acquiring one. It seems more natural to say that super-benevolence is a virtue, but one that happens not to be instantiated. Of course, this requires us to say that there can be uninstantiated properties; those who find this commitment problematic will be unmoved by the Uninstantiated Virtue Argument.

VI. CONTRASTIVIST COUNTERFACTUALIST VIRTUE CONSEQUENTIALISM

We are almost ready to state virtue consequentialism. But in order to state the view correctly, we must take a brief detour into the semantics of subjunctive conditionals.

According to David Lewis, a conditional of the form 'if it were A, then it would be that C' is non-vacuously true at world *w* if and only if at all of the closest (i.e. most similar) accessible A-worlds to *w*, C is true.[30] Lewis also holds a 'strong centring' view – no matter what respects of similarity are deemed relevant, no world W1 is as similar to another world W2 as W2 is to itself.[31] Jean-Paul Vessel has argued convincingly, for reasons I cannot fully explain here, that consequentialists should reject strong centring in the analysis of counterfactual and subjunctive conditionals.[32] Consider Lucky's world; call it W_L. The conditional 'if Lucky were malicious, he would not produce bad consequences' is usually thought to be true at W_L, given that both the antecedent and consequent are true at W_L. But that assumes strong centring; if other worlds can be as similar to W_L as W_L is to itself in the relevant respects, then there is no guarantee that the conditional turns out true. And in fact, given that it is a one-in-a-million chance that Lucky does not produce bad consequences, there are many worlds as similar in the relevant respects to W_L as W_L is to itself in which Lucky does produce bad consequences. These are worlds in which Lucky attempts his murders in slightly different ways from the way he actually attempts them, e.g. he aims a fraction of an inch to the left or right, or shoots a fraction of a second earlier or later, thus missing the bird and hitting

[30] Lewis, 'Counterfactual Dependence and Time's Arrow', *Conditionals*, ed. F. Jackson (Oxford, 1991), p. 56.
[31] Lewis, *Counterfactuals* (Oxford, 1973), p. 29.
[32] See Vessel, 'Counterfactuals for Consequentialists', *Philosophical Studies* 112 (2003).

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

his target. So the conditional turns out false. I will employ this view about similarity of worlds to account for Lucky.

Time to unveil the view. Incorporating contrastivist and counterfactualist positions into our view, and making use of the idea that a world W1 can be as similar to another world W2 as W2 is to itself, I propose we understand virtue consequentialism in the following way:

> VC: It is a virtue for people S1–Sn to have character trait V1 rather than character trait V2 at world *w* iff (i) V2 is a member of the contrast class of V1, and (ii) the expected intrinsic value of a closest world to *w* where S1–Sn exercise V1 is greater than the expected intrinsic value of a closest world to *w* where S1–Sn exercise V2.

Some features of VC require explanation. VC is a contrastivist view; it tells us under what conditions it is a virtue for some people at a world to have one character trait rather than another. Virtue is not an intrinsic property of character traits, but a quaternary relation that relates a set of people, two character traits, and a world. (We might also wish to consider whether virtues are also relative to periods of time, or to societies, resulting in a relation with five or six terms. I leave these questions aside.) VC accommodates both individualist and universalist views. We find what traits are virtues for an individual by plugging in just that individual for S1–Sn; we find what traits are virtues for everyone by plugging in everyone for S1–Sn.

'Expected intrinsic value' is a technical term, defined as follows. There are, naturally, infinitely many worlds that are most similar to a given world. Those worlds might not all have the same intrinsic value. Partition the set of closest worlds according to intrinsic value, so that all worlds with the same intrinsic value fall within the same partition. For each partition, there is a certain probability that a world from that partition would be actualized were the character trait in question exercised by the relevant people.[33] Multiply that probability by the intrinsic value of a world from that partition; do the same for all the other partitions; sum the products. The sum is what I am calling the 'expected' intrinsic value of a closest world. (I give it this name because of the obvious similarity to the notion of expected utility.)

To illustrate how this works, consider the case of Lucky again. Let W_L be the world where Lucky repeatedly fails to murder people owing to bizarre coincidences. One of the closest worlds to W_L in which Lucky

[33] I make no commitments about the sort of probability that should be employed here, e.g. 'subjective' or 'objective' probability. The sort of probability employed would determine whether VC entails what Driver calls 'evaluational internalism' or 'evaluational externalism' (*Uneasy Virtue*, p. xv).

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

exercises his malice is W_L itself. There are other worlds just as close to W_L in which Lucky is successful in his murder attempts. The reason why they are just as similar is that they are worlds where Lucky attempts his murders in slightly different ways – so similar that he has no ability to control which of the worlds is actualized. He can control whether he pulls the trigger today or tomorrow, or shoots left or right, but his control does not extend to tiny fractions of a second or fractions of an inch. To simplify matters, suppose the set of closest worlds where Lucky is malicious contains only two partitions: one contains worlds where Lucky succeeds; one contains worlds where he fails. All the worlds where he succeeds have much lower intrinsic values than worlds where he fails, and it is much more likely that Lucky succeeds than that he fails. Given plausible assumptions about what would happen if Lucky were not malicious, the expected intrinsic value of a closest world to W_L where Lucky is malicious turns out to be much lower than the expected intrinsic value of a closest world where he is not malicious.

To see how VC accounts for cases of unexercised virtue, recall the case of Helper. Let W2 be the world where Helper's generosity goes unexercised. A closest world to W2 where Helper's generosity is exercised is presumably a world where her exercising her generosity brings about good results. Her exercising some alternative trait (selfishness, say) would presumably bring about worse results. Thus, VC entails that it is a virtue for her to be generous rather than selfish at W2. Uninstantiated virtues like super-benevolence get treated in just the same way.

VII. COMPARING VIRTUES

Perhaps not all virtues are created equal. Driver suggests there is a hierarchy of virtues:

> The better virtues will be the ones that produce more good. Generosity is probably better than wit in that it produces more good. If there is a virtue that produces more good than any other, then that would be the best.[34]

It is not clear that we would ever need to make such claims, but suppose we do. Two interesting conclusions follow.

First, we have even more reason to reject actualism. Suppose there were just one generous person in the world, but billions of witty ones. The good consequences of generosity in the actual world would be limited; wit would have much better consequences, simply because of the sheer number of witty people. Each instance of wit might produce just a tiny bit of goodness, but owing to the billions of instances, the

[34] Driver, *Uneasy Virtue*, p. 74.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

total good produced by wit would be substantial. So on Driver's view, wit would turn out to be a better virtue than generosity in such a world. It seems wrong to say that wit is better than generosity simply in virtue of being more commonly exemplified.

Second, if we accept contrastivism about virtue, when we compare two character traits such that neither is in the contrast class of the other, such as generosity and wit, our comparison turns out to be complicated. When we say that a trait is a virtue, we are saying it is a virtue to have that trait rather than some other one. Thus, to say that one trait, V1, is a better virtue than another, V2, is to say that the extent to which it is a virtue to have V1 rather than V3 (where V3 is some trait from V1's contrast class) is greater than the extent to which it is a virtue to have V2 rather than V4 (where V4 is some trait from V2's contrast class). Whether such a claim is true depends on the choice of V3 and V4.

Since we are comparing V1 and V2 to different traits, the comparison between V1 and V2 might seem pointless. But it is not entirely pointless. I might want to become a more virtuous person, and to that end, I might want to know whether it would be better for me to acquire a certain amount of generosity or a certain amount of wit. Suppose it is equally in my power to do either, but I cannot do both. Let G1 and W1 be my current levels of generosity and wit, and G2 and W2 be the new, higher levels I might acquire. Suppose the difference in expected intrinsic value between a closest world where I have G1 and one where I have G2 is greater than the difference in expected intrinsic value between a closest world where I have W1 and one where I have W2. Then in so far as I care about being virtuous, G2 would be the better trait for me to acquire.

VIII. CONCLUSION

Nothing I have said constitutes an argument that VC is superior to Aristotelian or Kantian theories of virtue. (In fact, a version of VC, modified in obvious ways, could serve as part of the neo-Aristotelian view according to which virtues benefit their possessor.[35]) I have argued that VC is the best version of virtue consequentialism. VC is a contrastivist view, so it can account for the example of Downer. It can account for the example of Lucky through the machinery of expected intrinsic value and the rejection of strong centring. And as it is a counterfactualist view, it can account for the examples of Helper and super-benevolence. As VC accommodates individualism and

[35] See Hursthouse, *On Virtue Ethics*, p. 167, and Alasdair MacIntyre, *After Virtue* (Notre Dame, 1984), p. 191, for examples of such views.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press

universalism, it can account for our ambivalence about the courageous villain. In all these ways it is superior to extant virtue consequentialist theories. I hope that some might regard VC as a viable alternative to Aristotelian and Kantian views.[36]

wbradley@syr.edu

[36] I thank Fred Feldman, Andre Gallois, Eric Moore, Erik Wielenberg and an anonymous referee for helpful comments on earlier drafts. A previous, much shorter version of this article was presented at the 2002 Northwest Philosophy Conference. Thanks to all those present, especially Agnes Patkowski and Chenyang Li, for their comments.

https://doi.org/10.1017/S0953820805001652 Published online by Cambridge University Press