Decision-Theoretic Virtue Ethics

Ralph Wedgwood

All action is undertaken in the presence of uncertainty. In the face of such uncertainty, as Bishop Butler taught us long ago, "Probability is the very Guide of Life".[1] But how exactly should we be guided by probabilities in deciding how to act?

In the three centuries since Butler wrote, theorists have devised just one plausible answer to this question. In some sense, we should be guided by some kind of *probabilistic expectation* of some kind of *value*. This notion of "expected value" is the central insight of *decision theory*. In this paper, I shall focus on how this central insight of decision theory can be combined with ethical theory.

It is clear enough that decision theory can be married with *consequentialist* ethical theories.[2] However, many of us have ethical intuitions that sharply diverge from all standard forms of consequentialism. For this reason, we should examine how decision theory can be united with non-consequentialist ethical theories. Strangely, this important question remains seriously under-explored.

In what follows, I shall propose that the answer to this question can be found in some of the key insights of Joseph Raz – in particular, in his central principle that there is a fundamental connection between *reasons for action* and *values*. As he put it: "reasons [for actions] are facts in

---

[1] See Joseph Butler, *The Analogy of Religion* (London, 1736), Introduction, p. iii.

[2] See especially Frank Jackson, "Decision-theoretic consequentialism and the nearest and dearest objection", *Ethics* 101, no. 3 (1991): 461-482.

virtue of which those actions are good in some respect and to some degree".[3] Raz was always careful to formulate this principle in such a way that it did not involve any commitment to any form of consequentialism. In short, his view of reasons for action was *values-based* but *compatible with non-consequentialism*. As I shall show here, there is a plausible and coherent way to combine this view of reasons for action with a broadly decision-theoretic approach, which allows for the view to be developed in a resolutely non-consequentialist direction.[4]

More specifically, the version of this approach that I shall develop will be a form of *virtue ethics*. Admittedly, it is not what is today the best-known form of virtue ethics – the form that has been developed most prominently by Rosalind Hursthouse.[5] As I shall explain, it is closer to the form of virtue ethics that was entertained at one point by Judith Thomson; and in my view it is also closer to the form of virtue ethics that was articulated by Aristotle. My main goal here is to explain how this kind of virtue ethics can be combined with decision theory. IN

---

[3] See Raz, *Engaging Reason* (Oxford: Oxford University Press, 1999), Chap. 2 ("Agency, Reason, and the Good"), p. 23.

[4] At least one attempt has been made to unite decision theory with ethical theories that are "deontological", in the sense of not being values-based in the way that Raz's view is; see Seth Lazar, "Deontological Decision Theory and Agent-Centered Acts", *Ethics* 127, no. 3 (2017): 579-609. In my view, Lazar's account faces many grave problems, which can all be avoided by my Raz-inspired values-focused approach. Unfortunately, however, I cannot explain these problems with Lazar's account here.

[5] See Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999). For a sophisticated alternative to Hursthouse's approach, see Christine Swanton, *Target Centred Virtue Ethics* (Oxford: Oxford University Press, 2021); my view is certainly closer to Swanton's than to Hursthouse's, although it is not exactly the same as either.

short, the goal here is to articulate a form of *decision-theoretic virtue ethics* (DTVE).

In the first section, I shall canvas some cases that involve various forms of uncertainty – cases that any adequate ethical theory needs to give a convincing account of. In the second section, I shall explain how I understand the distinction between consequentialism and non-consequentialism; this will make it clear how Raz's values-focused view of reasons for action is compatible with rejecting consequentialism. In the third section, I shall give a sketch of the kind of virtue ethics that I shall assume here. Then, in the fourth section, I shall outline the kind of DTVE that seems to me most promising; and in the fifth section, I shall explain how this version of DTVE can provide intuitively plausible verdicts on the cases that I considered in the first section. My conclusion is that the considerations advanced here give us reason to think that this version of DTVE should be regarded as a serious contender in contemporary ethical theory.

1. Some problem cases

Let us start by canvasing the kinds of cases that an adequate theory needs to handle. In the first kind of case, there is a determinate fact of the matter about the relevant moral properties of the available acts, but the agent is not certain of this fact.

A good example of this kind of case is provided by Seth Lazar:[6]

> *Self-Defence*: Alice can defend herself from a lethal attack only by killing Bill and using his body to shield hers. Alice is almost certain that Bill, a business competitor, ordered the attack. Alice has no dependents and no outstanding obligations.

[6] Ibid., p. 587.

Intuitively, if Bill did order the attack, then he is "liable" to "defensive harm".[7] In this case, Alice would not be violating his rights in killing him and using his body as a shield to defend herself. On the other hand, if Bill did not order the attack, then he is not liable to this kind of defensive harm, and Alice would be violating his rights in killing him in order to defend herself.

We can make sense of a notion of "*objective* permissibility", on which the objective permissibility of Alice's killing Bill in this way depends on the objective facts about whether or not Bill actually ordered the attack. If Bill did in fact order the attack, Alice's killing him is "objectively permissible", whereas if Bill did not order the attack, Alice's killing him is not "objectively permissible". However, it seems that we can also make sense of a notion of "*subjective* permissibility", on which the subjective permissibility of Alice's killing Bill depends on the degree of confidence that it is rational for Alice, given the evidence that she has, to have in various hypotheses about these facts. If it is rational for Alice, given the evidence that she has, to have a very high degree of confidence in the hypothesis that Bill ordered the attack, it seems that it is "subjectively" permissible for her to kill Bill. If it is not rational for Alice, given her evidence, to have anything more than a low degree of confidence in the hypothesis that Bill ordered the attack, it is not subjectively permissible for her to kill Bill.

Suppose that it is rational for Alice, given her evidence, to have a high degree of confidence in the hypothesis that Bill ordered the attack, but in fact this hypothesis is false – Bill did not order the attack. In this case, Alice's killing Bill is objectively impermissible, but subjectively permissible. It is plausible that this implies that, in this case, Alice violates Bill's rights, but has an *excuse* for so doing – so that she is not blameworthy for her action, even

[7] For a cutting-edge discussion of the idea of "liability to defensive harm", see Jonathan Quong, *The Morality of Defensive Force* (Oxford: Oxford University Press, 2020), Chap. 2.

though it is objectively wrong. At all events, this example seems to show that whether an act counts as subjectively permissible can vary with the degree of confidence that it is rational for the agent, given her evidence, to have in various hypotheses about the relevant facts of the case.

As Douglas Portmore has pointed out, however, parallel issues arise about objective permissibility, in cases where there is *no determinate fact of the matter* about what would have happened had the agent performed one of the acts that she did not actually perform. One way in which this can happen is if the causal structure of the world is *indeterministic*. For an example of this kind of case, consider this variant of Portmore's case of the "Questionable Man":[8]

> Abaddon has been told that a terrorist group has kidnapped his brother, and that his brother will be killed unless Abbadon detonates a bomb in the marketplace, killing 20 people. (In fact, this is false: Abbadon's brother is in no danger, but Abbadon gives a high degree of credence to what he has been told.) Shamira has the option of killing Abaddon at a certain particular point in time – before he has an opportunity to detonate the bomb. However, the workings of Abaddon's mind are *indeterministic*. It is not true that, if he were not killed, he would detonate the bomb, nor is it true that, if he were not killed, he would not detonate the bomb. All that is true is that, if he is not killed, there is a certain objective *chance* that he will detonate the bomb, and a certain objective *chance* that he will not. Shamira's only other option is to refrain from intervening, allowing Abaddon to do whatever he finally decides to do.

If Shamira kills Abbadon, there is no determinate fact of the matter about what he would have

---

[8] See D. W. Portmore, *Opting for the Best: Oughts and Acts* (Oxford: Oxford University Press, 2019), pp. 239f.

done had she not killed him. When his mind is in this indeterministic state, Abbadon seems a borderline case of being "liable to defensive harm". So, it is objectively permissible for Shamira to kill him or not? It seems, intuitively, that this depends on the objective *chances* of his detonating the bomb if he is not killed. If the chance that he will detonate the bomb if he is not killed is extremely *high*, it seems objectively permissible for Shamira to kill him; if the chance that he will detonate the bomb if he is not killed is extremely *low*, it seems not to be objectively permissible for her to kill him. In this way, even the factors that determine *objective* permissibility seem susceptible to being affected by the objective chances.

Finally, as Portmore also shows, there is a third kind of case that could in principle arise even if the causal workings of the world are fundamentally deterministic. This third kind of case can arise if the principle that Krister Bykvist has called "*counterfactual* determinism" is false.[9] According to counterfactual determinism, for every act that is available to an agent at a time, there is a *unique possible world* that would obtain if the agent were to perform that act at that time. We can illustrate this principle with the following example, which focuses on your situation as you aim a dart at the dartboard: according to this principle, either it is the case that, if

[9] See Bykvist, "Normative Supervenience and Consequentialism", *Utilitas* 15, no. 1 (2003): 27–49, p. 30. Under natural assumptions, counterfactual determinism follows from the principle that philosophical logicians call "Conditional Excluded Middle". This is the principle that implies every instance of the following schema 'Either if it were the case that *p* it would be the case that *q*, or if it were the case that *p* it would *not* be the case that *q*'. This principle was defended by Robert Stalnaker and denied by David Lewis. See Lewis, *Counterfactuals* (Oxford: Blackwell, 1973), and Stalnaker, "A defense of conditional excluded middle", in William Harper, Robert Stalnaker, and Glenn Pearce (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time* (Dordrecht, Holland: D. Reidel, 1981), 87–104.

you were to try to hit the bullseye, you would succeed, or it is the case that, if you were to try to hit the bullseye, you would not succeed. By contrast, if the principle is false, then – at least if you do not *actually* try to hit the bullseye – it may not be the case either that if you were to try you would succeed, or that if you were to try you would not succeed. At most, there is a certain conditional *chance* of your succeeding, conditionally on your trying, and also a conditional chance of your not succeeding, conditionally on your trying.

The relevance of this issue about counterfactual determinism may be illustrated by the following example, which is also adapted from a case that is due to Portmore:[10]

> Ka'eo is at the dartboard, when an evil demon presents him with the following dilemma. If Ka'eo stops playing darts immediately, the demon will do nothing. If Ka'eo throws the dart and does not succeed in hitting the bullseye, the demon will kill an innocent person. If Ka'eo throws the dart and succeeds in hitting the bullseye, the demon will give each of 10,000 people a small reward. Presented with this dilemma, Ka'eo stops playing darts, and does not try to hit the bullseye.

Suppose that counterfactual determinism is false, and so it is not true either that if Ka'eo had thrown the dart he would have succeeded in hitting the bullseye, or that if he had thrown the dart he would not have succeeded in hitting it. There are merely certain conditional objective chances of his succeeding, and of his not succeeding, conditionally on his throwing the dart. Intuitively, if the conditional chances of success are extremely high, it is objectively permissible for him to throw the dart, but if these chances are low, it is not objectively permissible for him to throw the dart. Again, in this case, these conditional objective chances seem to make a difference to the

[10] Portmore, ibid., p. 242.

objective permissibility of the available acts.

These are challenging problems to solve. To see how challenging they are, it will help to consider one *prima facie* promising approach, which in fact fails to solve all these problems satisfactorily. It might seem that in a sense the objective wrongness of acts *comes in degrees*: some acts are only *slightly* wrong, while other acts are *very* wrong, and so on. Suppose that we can actually *measure* the degree to which acts are wrong. If we can indeed measure an act's degree of wrongness, then we may be able to make sense of an act's *expected degree of wrongness*, in terms of some appropriate probability function. Perhaps, then, we could say that for an act to be subjectively permissible, its expected degree of objective wrongness (in terms of the relevant probability function) must be no greater than that of any alternative: in short, subjectively permissible acts are those that *minimize expected wrongness*.[11]

While this approach sounds promising, it cannot yield the right solution in cases like Lazar's case of *Self-Defence*. If Alice refrains from killing Bill and using his body as a shield, then – regardless of whether Bill ordered the attack or not – it seems that her act is *certain* not to be objectively wrong to *any degree* at all: indeed, if it in fact it is not wrong for Alice to kill Bill, then Alice's sacrificing her life in this way seem to be an act of saintly supererogation. On the other hand, if she kills Bill and uses his body as a shield, then it seems that there is some non-zero probability that her act is objectively wrong. Thus, if Alice sacrifices her life, she is certain not to be acting wrongly to any degree, while if she kills Bill and uses his body as a shield, there is some non-zero probability that her act is seriously wrong. It follows that the only way for her

[11] For the idea that moral wrongness comes in degrees, and a morally conscientious agent seeks to *minimize* the wrongness of her acts (in the light of the information that she possesses), see Peter A. Graham, "In Defense of Objectivism about Moral Obligation", *Ethics* 121, no. 1 (2010): 88–115, p. 99.

to minimize expected wrongness is by sacrificing her life. So, proponents of this approach will have to say that the *only* subjectively permissible act for Alice in this case is to sacrifice her own life.

As Lazar convincingly argues, this is an extreme view, which most of us who are not radical pacifists are strongly inclined to reject.[12] It implies that self-defence is never subjectively permissible in the real world, since whenever we engage in self-defence in the real world, we are acting in the face of significant uncertainty about some of the relevant facts. Intuitively, if Alice is rational, given her evidence, in being virtually certain that Bill ordered the attack on her life, it is "subjectively" permissible for her to kill him in self-defence. An adequate theory will have to explain why this act is subjectively permissible in the case.

2. Consequentialism and non-consequentialism

As I shall understand it here, consequentialism about acts is the view that the ethical status of acts is fixed by some kind of *value* of the *consequences* of those acts – where a "consequence" of an act is, in effect, a *whole possible world* that could result from the act's being performed. In other words, for consequentialists it is the value of whole possible worlds that is fundamental; the ethical properties of acts are fixed by their relations to these possible worlds.

This kind of consequentialism has a simple and elegant way of explaining the way in which probability affects the ethical status of acts.[13] In general, consequentialism of this sort

[12] Lazar, ibid., p. 588.

[13] In fact, it is doubtful whether consequentialism, at least as I formulate it in this section, can yield the right verdict in cases, like Seth Lazar's *Self-Defense* case, in which some of the available acts

needs just two elements: first, some way of measuring the relevant value of worlds; and, secondly, some appropriate probability function. These consequentialists could then explain the difference between objective permissibility and subjective permissibility in terms of this probability function. For objective permissibility, the relevant probability function is an *objective chance function*; for subjective permissibility, the relevant probability function is some kind of *evidential* or *subjective* probability.[14] Let $V(\bullet)$ be the relevant measure of the value of worlds, and $P(\bullet)$ the relevant probability function. Then, for every act A, we can calculate A's *expectation* – a probability-weighted sum of the values of the worlds compatible with A, weighting each of these values by the relevant conditional probability of the world, conditional on A's being performed:

$$\sum_w V(w)\, P(w|A).$$

The available acts can then be *ranked* in terms of their expectations. According to the consequentialist, an act is permissible (in the sense corresponding to the probability function $P(\bullet)$) if and only if there is no alternative act that has a higher expectation.

Most consequentialists assumed that the relevant value of these worlds is a purely *agent-neutral* and *time-neutral* value – a value that does not treat any particular agent or time as special, but is in a way neutral between all agents and times whatsoever. For example, the

---

are *supererogatory*. What is important here is just that consequentialism has no problem in explaining how probability affects permissibility.

[14] For the distinction between interpreting a probability function as an objective chance function and as an evidential or subjective probability, see Alan Hájek, "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed. Edward N. Zalta <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>.

classical utilitarians supposed that the relevant value of each of these worlds was simply given by the total quantity of happiness in that world as a whole. Importantly, however, as Portmore and others have shown, this consequentialist treatment of these cases is compatible with allowing that the relevant value of the worlds is not agent-neutral, or even time-neutral, but instead either agent-relative or time-relative or both.[15] Relative to the situation that *I* am in now, a world w_1 where *I* kill one person now, and you fail to prevent me, may have less value than a world w_2 that differs from w_1 only in that you and I switch places – that is, a world where *you* kill one person now, and I fail to prevent you. Moreover, relative to the situation that I am in *now*, worlds where a *past* act of mine turns out to kill someone, but no present act of mine kills anyone, may have greater value than worlds where no past of mine kills anyone, but one of my *present* acts turns out to kill someone. Even if the relevant value of worlds is this kind of agent-relative and time-relative value, the consequentialist formula can still be deployed to explain which acts are permissible in the relevant sense – that is, objectively permissible if the probability function is an objective chance function, and subjectively permissible if it is some kind of subjective or evidential probability function.

The key point about consequentialism – even of this agent-relative and time-relative variety – is that it implies that the permissibility of acts is entirely *derivative* from the *value* of *worlds* (together with the relevant probability function). Each of these worlds may be thought of as a "possible total consequence" of the relevant act. So, another way to put the point is by pointing out that, according to consequentialism, the permissibility of each act is entirely

---

[15] See Portmore, ibid., pp. 235f. The idea of such agent-relative forms of consequentialism was originally due to Amartya Sen, "Evaluator Relativity and Consequential Evaluation", *Philosophy and Public Affairs* 12, no. 2 (1983): 113-132.

derivative from the values of all the *possible total consequences* of the available acts.

If consequentialism is incorrect, then the permissibility of each act is not wholly derivative from the values of the worlds that count as the possible total consequences of the available acts. The permissibility of the act is determined by some other factor instead. This does not mean that non-consequentialists must deny that this other factor involves values of any kind; they must simply deny that this other factor is limited to the values that are exemplified by the possible total consequences of the available acts – that is, to values that are exemplified by whole possible worlds.

This point was quite clear to Joseph Raz, who consistently insisted that every reason for an action is "a fact in virtue of which the *action* is good" – and carefully avoided committing himself to the consequentialist view that the value of acts is always derivative from the value of the *consequences* of those acts.[16] In short, for Raz, it is quite possible that one act *A* is better in the relevant way than an alternative act *B* even if the total consequences of *A* are not in any relevant way better than the consequences of *B*. In such a case, there is presumably more reason, all things considered (ATC), for the agent to choose *A* than to choose *B*, even though the consequences of *A* are not in any relevant way better than the consequences of *B*.

There are many ways in which this non-consequentialist conception of the value of acts could be developed. For my part, I am sympathetic to a theory that incorporates two traditional distinctions – the distinction between *doing* and *allowing*, and the distinction between *intended*

[16] See the formulations that Raz gives in *Engaging Reason*, pp. 1, 23, 47, 97, and many other places.

and *unintended* effects that is central to the Doctrine of Double Effect.[17] On this picture, the value of an act is determined partly by the value of each of its significant effects, but also by the *relation* that the act has to that effect – for example, by whether it has the *doing* or the *allowing* relation to the effect, or whether or not the act involves the *intention* to bring about the effect. Thus, an act of killing a person can be worse than an act of letting a person die, even if the total consequences of both acts – the whole possible worlds that result from those acts – do not differ with respect to any of the relevant values. For our present purposes, however, we do not need to inquire into these details about the values of acts. All that matters is that we (a) make it clear that appealing to these values does not commit us to any form of consequentialism, and (b) explain how these values can interact with probabilities or chances in the way that the examples that we considered in Section 1 seem to require.

In this discussion, I shall investigate the prospects of a non-consequentialist theory that is in a sense a form of *virtue ethics*. However, I shall understand virtue ethics more broadly than is common in many contemporary debates. Many varieties of contemporary virtue ethics interpret the virtues primarily as *character traits* of *individual agents*.[18] However, as I shall now argue, this is not the only possible form that virtue ethics can take.[19]

Importantly, the virtue words – words like 'just', 'generous', 'prudent', 'courageous', and

---

[17] See my earlier works "Intrinsic Values and Reasons for Action", *Philosophical Issues* 19 (2009): 342–363, and "Defending Double Effect", *Ratio* 24, Special Issue: Deontological Ethics, ed. Brad Hooker (2011): 384–401.

[18] See especially Hursthouse, *On Virtue Ethics*, p. 28.

[19] Indeed, in my view, Hursthouse's version of virtue ethics faces grave objections – a number of which are laid out by Robert N. Johnson, "Virtue and Right", *Ethics* 113 (2003): 810–834.

the like – are capable of describing many other phenomena besides individual agents. In particular, as Judith Thomson has noted, the items that can be called "just" or "unjust", "generous" or "ungenerous", "prudent" or "imprudent", clearly include *acts* as well as *agents*.[20] As Thomson also argued, these words refer to "ways of being good": an act that is just or generous is good in a way, while an act that is unjust or miserly is bad in a way.[21] Moreover, these forms of goodness are not primarily instantiated by worlds or consequences: they are instantiated by the acts themselves. If a theory implies that what it is permissible for an agent to do at a time is determined by the various virtues and vices that are instantiated by the *acts* that are available to that agent at that time, that will also count as a form of virtue ethics by my lights. I shall call such theories *act-focused* versions of virtue ethics.

These act-focused versions of virtue ethics can clearly be developed in a way that allow them to diverge from all agent-neutral and time-neutral forms of consequentialism. The simplest way to show this is to focus on *agent-centred constraints*. In general, agent-cantered constraints are incompatible with agent-neutral forms of consequentialism, since these constraints imply that it can be permissible for you to do *A* and impermissible for you to do *B* even if the total consequences of *A* are no better than those of *B* in terms of all agent-neutral values.[22] For

---

[20] See J. J. Thomson, *Goodness and Advice*, ed. Amy Gutmann (Princeton, New Jersey: Princeton University Press), Part Two, §5 (pp. 58–65).

[21] Thomson, ibid., Part Two, §6 (pp. 65–67).

[22] If, as Bykvist has argued in "Normative Supervenience and Consequentialism" (see n. 9 above), consequentialism is fundamentally the thesis that the permissibility of acts *supervenes* on the value of the acts' total consequences, this is the cleanest kind of counterexample to consequentialism – a

example, suppose that if you do *A*, I will kill an innocent person – Person 1 – thereby preventing you from killing a second innocent person – Person 2 – while if you do *B*, you will kill Person 2, thereby preventing me from killing Person 1. We may suppose that the consequences of your doing *A* and of your doing *B* are exactly alike in all ethically relevant respects, except that two pairs of individuals (you and I, and Person 1 and Person 2) have switched places. Thus, in the consequence of *A*, you neither kill nor save anyone, while I kill Person 1 and save Person 2, while in the consequence of *B*, I neither kill nor save anyone, while you kill Person 2 and save Person 1. Since these consequences differ only in this permutation of individuals, the consequences of *A* and *B* are exactly as good in terms of all agent-neutral values. Nonetheless, a theory that accepts agent-centred constraints might imply that in this case, it is permissible for you to do *A*, but not *B*.

An act-focused version of virtue ethics can explain such agent-centred constraints in the following way. If you do *A*, you fail to act *beneficently* towards Person 1, but also do not act *unjustly* towards anyone, since you do not yourself kill anyone or violate anyone's rights. By contrast, if you do *B*, you act beneficently towards Person 1, but you also act unjustly towards Person 2, killing them and thereby violating their rights. Let us assume that in consequence there is a reason of beneficence against *A* and in favour of *B*, and a reason of justice against *B* and in favour of *A*.[23] If this reason of justice outweighs this reason of beneficence, then all things

---

case where two acts differ in their permissibility even though the value of their total consequences is the same.

[23] Many virtue ethicists have argued that in these cases the reason of justice somehow *cancels* or *disables* any reason of beneficence; for views along these lines, see Thomson, ibid., p. 63, and Philippa Foot, "Utilitarianism and the virtues", *Mind* 94, no. 374 (1985): 196-209. In the model that I shall develop

considered (ATC) there is more reason for you to do *A* than to do *B*. At least so long as these are the only two acts available to you, this may explain why it is permissible for you to do *A* but not to do *B*. As the virtue ethicist might say, in this case your doing *A* is overall the most virtuous thing for you to do, while your doing *B* is overall less virtuous than doing *A*. For the virtue ethicist, if an act is in this way overall the most virtuous thing to do, it is also what ATC there is most reason to do, and is thereby permissible.

It is less obvious that this kind of virtue ethics differs from agent-relative and time-relative versions of consequentialism, since it is widely believed that these relative forms of consequentialism can extensionally agree with any other ethical theory in which acts they classify as permissible and which they classify as impermissible.[24] However, even if there is an agent-relative and time-relative form of consequentialism that agrees *extensionally* with this kind of virtue ethics, the two theories still disagree in the *explanation* that they give of why these acts

---

below, the reason of beneficence is not cancelled or disabled in these cases, but simply outweighed. Unfortunately, I shall not be able to defend this approach here, although I believe that I could do so by redeploying an argument that I developed in "The Weight of Moral Reasons", *Oxford Studies in Normative Ethics* Vol. 3, ed. Mark Timmons (Oxford: Oxford University Press, 2013): 35–58. This is not to say that such cancelling never occurs: it may be that the fact that a judge's opinion would be *wittier* if she rules in favour of the prosecution simply fails to provide any reason for the judge to rule one way than another. For an explanation of how this kind of cancelling is compatible with my approach, see "The Reasons Aggregation Theorem", *Oxford Studies in Normative Ethics*, Vol. 12, ed. Mark Timmons (Oxford University Press, 2022), 127–148, p. 146.

[24] See Portmore, *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press, 2011), Chap. 4.

are permissible. The consequentialist theory gives explanatory primacy to the value of the possible worlds that count as the possible total consequences of the available acts. The act-focused forms of virtue ethics disagree with this: in their view, what has explanatory primacy is not the value of these worlds or consequences, but the virtue-properties of the acts themselves.

In this way, it is clear, at least in broad outline, how this act-focused version of virtue ethics differs from consequentialism. It differs from consequentialism because the kind of goodness and badness to which it fundamentally appeals is not the goodness and badness of whole worlds or total consequences, but the goodness and badness of acts themselves.

The crucial feature of these virtue-properties for our purposes is that – since they are ways of being good, as Thomson puts it – they *come in degrees*. Some unjust acts are *more unjust* than others; some beneficent acts are *more beneficent* than others; and so on. In this way, a form of virtue ethics that appeals to these properties of acts makes room for an essentially *scalar* theory – a theory on which the relevant properties may be possessed by different acts to different degrees, allowing for continuous variation from one case to another. This is the most fundamental way in which both consequentialism and virtue ethics differ from the dominant forms of deontology. These dominant forms of deontology are fundamentally *non-scalar*: they seek to draw a precise line separating right acts from wrong acts, without appealing to any normative or evaluative property that comes in degrees.[25] As I shall try to show, it is the fact that virtue ethics is a scalar theory in this way that makes it possible to unite virtue ethics with decision theory.

---

[25] For example, according to Kant, at least as I read him, acting on a maxim is permissible if and only if the maxim is in the relevant sense "universalizable" – where universalizability is not a feature of maxims that comes in degrees; see Kant, *Groundwork of the Metaphysics of Morals* (4: 401, 421).

3. Assumptions about the virtues

One question that a complete version of virtue ethics would have to answer concerns *which* virtues are relevant to the permissibility of acts. Unfortunately, I shall not be able to give a full answer this question here. In this section, I shall explain which issues about the virtues I can remain neutral about, and which assumptions about the virtues I shall need to rely on.

One crucial difference between virtues is the difference that we could articulate by distinguishing between "abstract virtue" and "manifesting virtuous dispositions".[26] As Aristotle pointed out in the *Nicomachean Ethics* (1105a17–b9), an agent might perform an act of type *A* at the same time as being in a situation in which it is *just* for her to perform an act of type *A* – even if it is simply a lucky fluke that she does something that it is just for her to do (perhaps it just so happens that her wicked plans require her to perform an act of type *A* in this situation). In this case, in Aristotle's terminology, the agent might be doing a *just act*, but she is not *acting justly*. In my terminology, the sense in which this act is a "just act" can be expressed by saying that it is "abstractly just" or "abstractly virtuous", while the sense in which it is not a case of the agent's "acting justly" can be expressed by saying that it is not a "manifestation" of "dispositions" that even partially constitute the agent's having the character trait of being – at least to some degree – a just person. In short, both abstract virtue and the manifestation of virtuous dispositions are good features of acts: but they differ in that an act can instantiate an abstract virtue through sheer luck, while if an act is the manifestation of a virtuous disposition, it is in a way no accident that the agent performs an action that has this feature.

---

[26] For this terminology, see my book *The Value of Rationality* (Oxford: Oxford University Press, 2017), Chap. 6.

Corresponding to this distinction, there are three different ways in which virtue ethics could seek to explain the facts about permissibility. First, some versions of virtue ethics could seek to explain these facts by appealing only to the *abstract virtues* that acts can instantiate.[27] Secondly, some versions could appeal only to the *virtuous* or *vicious dispositions* of the agent that are manifested in the relevant acts. Finally, some versions could appeal to *both* kinds of features of acts. In fact, I am inclined to favour the first of these three versions, but for the purposes of the present discussion, I can remain neutral on this question here.

A second way in which these virtue-properties of acts may differ is that some are in a way *particularized*, while others are not. One virtue-property that an act can have is the property of being *fair to Alfred* – or in other words, the property of treating Alfred fairly. Another such virtue-property is that of being fair *simpliciter* – which presumably is, in effect, the property of treating everyone fairly. The first virtue-property – being fair to Alfred – is a particularized virtue, while the second – being fair *simpliciter* – is not. In this way, we can make sense of particularized virtue-properties of many other kinds. An act can be beneficent *to Betsy* – in the sense of helping or benefiting Betsy; it can be just *to Carlos* – in the sense of treating Carlos justly, by respecting his rights rather than infringing them; and so on.

Again, there can be different versions of virtue ethics depending on which virtue-properties are given explanatory primacy in explaining what is permissible and what there is most reason for the agent to do. Some versions will appeal to particularized properties, while other will only appeal to non-particularized properties. Again, I shall remain neutral about which

---

[27] Thomson's version of virtue ethics is of this kind, because she insists on interpreting the relevant kind of "justice" and "generosity" as determined purely by the external causal features of the acts in question, and not by any facts about the agent's psychology; see Thomson, ibid., pp. 60ff.

version of virtue ethics is most promising here. However, to keep things simple, I shall not mention the particularized virtues in what follows: I shall write as though the non-particularized virtues are the only ones that matter for our purposes.

In general, I need not assume any particular list of the relevant virtues here. However, to fix ideas, it might be helpful to think of this approach as adopting something like W. D. Ross's list of *prima facie* duties – fidelity, reparation, gratitude, justice, beneficence, self-improvement, and non-maleficence – which we may reinterpret as a list of the relevant virtues.[28]

However, there are some assumptions about the virtues that are relevant to determining what is permissible that I shall need to rely on here. One particularly crucial assumption was already implicit in my explanation of how virtue ethics can explain agent-centred constraints. This is the assumption that there is a *plurality* of *potentially conflicting virtues* – for example, perhaps the virtues of justice and beneficence may conflict with each other in certain cases – and there is also, at least sometimes, some way of *aggregating* these conflicting virtues into an ATC judgment of how virtuous the available acts are overall.

To say that two virtues "conflict" is to say that these two virtues disagree in their ranking of the available acts. For example, perhaps in terms of the virtue of beneficence, one act *A* is better than a second act *B*, while in terms of another virtue, such as justice, *B* is better than *A*. Nonetheless, practical reason is not inevitably defeated whenever such conflicts between virtues arise. On the contrary, in many cases, it is possible to aggregate the verdicts of these different virtues together to produce a judgment of which act is most virtuous overall – or in other words, which act there is most reason ATC for the agent to do. For example, in the explanation that I sketched above of how virtue ethics can make sense of agent-cantered constraints, I proposed

[28] See W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930), p. 21.

that, in the case that I was focusing on, the "reason of justice outweighs the reason of beneficence". What this means more precisely, within my virtue-ethical framework, is that when the verdicts on the relevant acts of both justice and beneficence are aggregated, the act favoured by justice emerges as the most virtuous overall – which is what for the virtue ethicist makes it the act that there is, ATC, most reason to perform.

A further crucial assumption that I shall rely on is that the relevant virtues are not restricted to the *moral* virtues – at least not in the narrow sense of 'moral' that has to do with what we *owe* to *other people*. Most moralists in the history of philosophy have not restricted the virtues to the moral virtues in this narrow sense. Aristotle's virtues prominently include the intellectual virtues – such as technical skill, practical wisdom, and theoretical wisdom; and Raz himself was sceptical of the very distinction between moral and non-moral values.[29] Specifically, I shall assume from now on that the relevant virtues include *prudence* – where by 'prudence' I mean what the 18th-century British moralists called the virtue of *self-love*. For almost all the history of moral philosophy, the idea that rational self-love is a virtue was all but universally accepted. Even if rational self-love is not *morally* praiseworthy, the moral philosophers of the past seem to me correct in arguing that it is praiseworthy in a larger sense.[30]

However, the most important assumption about the virtues that I need here is that the virtues do not only come in degrees, but are also susceptible to a modest kind of *cardinal*

---

[29] See Aristotle, *Nicomachean Ethics*, Book VI, Chaps. 3–6 (1139b15–1141a7), and Raz, *Engaging Reason*, Chap. 11 ("The Moral Point of View").

[30] For a powerful argument that "true self-love" is one of the "heads" or "branches" of virtue, and that its manifestations are often praiseworthy, see Richard Price, *Review of the Principal Questions in Morals*, ed. D. D. Raphael (Oxford: Oxford University Press, 1974), p. 149f.

measurement. I cannot offer a full defence of this assumption here, but I can point to some intuitive considerations that seem to make it plausible. In particular, it seems that we can not only *compare* the available acts as more or less just, but we can also qualify these comparisons by saying that one act is *much* less just than another, but only *slightly* less just than a third, and so on. This seems to indicate that we can compare the *differences* or *intervals* between the degrees to which acts are virtuous: the difference between the degrees of justice of act A_1 and act A_2 might be a *smaller* difference than between the degrees of justice of A_2 and A_3. If these kinds of comparisons between differences are indeed possible, then it seems that some form of cardinal measurement of these degrees of justice should be possible.[31]

More precisely, I shall assume that these degrees of justice are capable of being measured on a so-called *interval scale*. In this respect, my assumption about degrees of justice resembles the classical decision theorists' view of *preferences* – since these decision theorists believe that it is possible to measure the strength of an individual's preferences for various prospects on an interval scale. Famously, the decision theorists' name for the function that measures the strength of an individual's preferences is the individual's "utility function". So, I am assuming here that we can *measure* acts' degrees of justice on an interval scale, by means of a function that will look formally a lot like a utility function – $V_j(\bullet)$.

In spite of the formal similarities between this measure of degrees of justice $V_j(\bullet)$ and a decision-theoretic utility function, there are some important differences. In particular, $V_j(\bullet)$ is a

[31] For the connection between the comparability of differences and measurability on an interval scale, see David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky, *Foundations of Measurement – Volume 1: Additive and Polynomial Representations* (New York and London: Academic Press, 1971), 150–2 and 157–8.

measure of the degree to which *acts themselves* are just or unjust. It is *not* a measure of the degree to which the *worlds* that are the acts' possible total consequences are good or desirable or whatever; this is how this approach is compatible with rejecting the consequentialist view. In this way, $V_j(\bullet)$ differs from utility functions – on the dominant way in which utility functions are interpreted by philosophers today – which are defined over possible worlds and propositions that correspond to sets of possible worlds.[32]

Still, so long as the same act can be performed in different possible worlds, an act's (objective) degree of justice may *vary from one world to another*. For example, Alice's act of killing Bill has a much higher degree of justice in the worlds where Bill ordered the attack on Alice than in worlds where he did not. So, this measure of degrees of justice must assign a value to each act *A relative to a possible world w* – $V_j(A, w)$. In what follows, I shall assume that all the virtues involved in determining which acts are permissible are measurable in this way: we can not only measure *A*'s degree of justice, relative to a world *w*, but also *A*'s degree of beneficence relative to *w*, *A*'s degree of prudence relative to *w*, and so on.

4. A precise version of DTVE

In the previous section, I explained my assumption that an act's degree of justice, relative to a possible world, can be measured on an interval scale. This will enable us to make sense of an act's *expected degree of justice* – and an act's expected degrees of *beneficence* and *prudence*, and so on – in terms of *any* probability function that is defined over the relevant space of worlds.

[32] For this way of understanding utility functions, see James M. Joyce, *Foundations of Causal Decision Theory* (Cambridge: Cambridge University Press, 1999), Chap. 4.

Let P(•) be a probability function defined over the relevant space of possible worlds. This probability function may be an objective chance function, or an evidential or subjective probability function; it does not matter – the same account will apply. In general, the degree of justice that an act A has at each world w can be *weighted* by the conditional probability of the world, conditional on the act's being performed – $V_j(A, w)\ \mathrm{P}(w \mid A)$. This determines the probability-weighted sum of these different degrees of justice that the act has at these different possible worlds –

$$\sum_w V_j(A, w)\ \mathrm{P}(w \mid A).$$

This is the act's expected degree of justice (relative to this probability function P(•)).

Now, as I explained in the previous section, to determine what there is *most reason* to do ATC, the reasons provided by these different conflicting virtues (V_1, … V_n – justice, beneficence, prudence …) need to be *aggregated* somehow. To keep things simple, I shall assume here that this aggregation will necessarily be *additive*: the expected degree to which the act is virtuous overall – that is, for the virtue ethicist, the expected degree to which there is reason ATC to perform the act – is a *weighted sum* of the expected degrees to which the act exemplifies the relevant virtues.[33]

To make sense of this, we need first to *normalize* our measurements of the degrees to which the acts exemplify each relevant virtue at each possible world. The simplest way to do this is, effectively, to measure degrees of *vices* rather than degrees of virtues. The available acts that are *optimal* in terms of each virtue do not exemplify the corresponding vice to any degree at all, and so their degree of vice is 0. Then we can measure the degree to which each of the other acts

[33] For a defence of this additive view of how to aggregate the different virtues, see my paper "The Reasons Aggregation Theorem" (cited in n. 23 above).

falls short of this optimal level of virtue, giving us the degree to which the act exemplifies the corresponding vice.

Then, for each relevant vice, we need to *weight* the measure V_i of that vice by a *factor* α_i that captures this measure V_i's importance for the agent's decision-making. An act A's expected overall degree of vice $OV(A)$, then, is:

$$\sum_i \sum_w \alpha_i \, V_i(A, w) \, P(w \mid A).$$

This idea of "weighting" these measures of the different virtues and vices seems to be implicit in Raz's idea that we can distinguish between "small" values and "big" values.[34] What Raz calls the "small values" are those that typically get weighted less heavily, and so rarely outweigh other values that conflict with them (in the sense that of making the act that exemplifies the small value the most virtuous act overall).

The acts with the *lowest* expected overall degree of vice are those with the *highest* expected overall degree of virtue. If the relevant probability function is an objective chance function, then these acts are those that the agent has *objectively* most reason to perform; if the relevant probability is an evidential or subjective probability function, these acts are those that the agent has *subjectively* most reason to perform.

Now, in a weak sense, we can allow that the different virtues are "incommensurable" – different agents may quite permissibly assign different weights ($\alpha_1, \ldots \alpha_n$) to these measures of different virtues. In particular, as Raz himself suggests,[35] a *supererogatory* agent may reasonably put less weight on the self-regarding virtue of prudence, and more weight on moral virtues such as justice or beneficence, than other less supererogatory agents would do. To fix ideas, we may

---

[34] See Raz, *Engaging Reason*, Chap. 2 ("Agency, Reason, and the Good"), p. 30.

[35] See Raz, *Engaging Reason*, Chap. 10 ("The Truth in Particularism"), Section 5, esp. p. 243.

assume that for each of the virtues, there are upper and lower bounds to the weights that agents may reasonably put on that virtue: neither totally neglecting either justice or prudence, nor giving infinitely greater weight to one virtue than to all the others, will be reasonable. But within these limits, we may suppose, there is a range of different weightings of the virtues, none of which is more reasonable than any other.

For an act to be *morally wrong*, I propose, is for the act to be *worse* in terms of the expected moral virtues (relative to the relevant probability distribution) than *every* act that is no less virtuous overall than any alternative on *any* reasonable weighting of the virtues. This implies that an act cannot be morally wrong unless it is less virtuous overall than some available alternative act on *every* reasonable weighting of the virtues – and moreover, it must be less virtuous overall than this alternative with respect to the *moral* virtues.[36] In practice, this means that we need to consider the reasonable weighting of the virtues that assigns the *greatest* weight to the non-moral virtues, and the *lowest* weight to the moral virtues; for the act to be wrong is for it to be morally worse than the maximally virtuous acts even on this minimally-moral weighting of the virtues.

If an act is morally wrong according to a subjective or evidential probability function that corresponds to the credences that it is rational for the agent, given her evidence, to have, then the

[36] This interpretation of moral wrongness makes room for the phenomenon that Elizabeth Harman has categorized as "permissible moral mistakes". In my framework, these are acts that for moral reasons one ought not to do, but which are not morally wrong, because they are still better in terms of the moral virtues than some alternatives that one has ATC most reason to do on *some* reasonable weighting of the virtues – namely, a weighting that attaches great weight to some non-moral virtues such as prudence. See Harman, "Morally Permissible Moral Mistakes", *Ethics* 126, no. 2 (2016): 366–93.

act is *subjectively* morally wrong (and, I would say, morally blameworthy). If the act is morally wrong according to an objective chance function, then the act is *objectively* morally wrong – though if it is not subjectively morally wrong, the agent's act is excusable and not blameworthy.

Finally, we can offer an account of what it is for acts to be morally permissible. An act is objectively morally permissible if and only if it is not objectively morally wrong; an act is subjectively morally permissible if and only if it is not subjectively morally wrong. This is how the version of DTVE developed here can explain what it is morally permissible for agents to do.

5. Solving the problem cases

The precise version of DTVE outlined in the previous section can be used to address the problem cases that we canvased in Section 1. We shall start with Seth Lazar's "Self-Defence" case. In my analysis, there are effectively two cases here, depending on whether Alice is a saintly hero or not. P(•) is a subjective probability corresponding to the credences that it is rational for Alice, given her evidence, to have.

Case 1: Alice is not a saintly hero

Act 1: Alice kills Bill and uses his body as a shield to protect herself.

In World W_1, Bill did not order the attack, and so is not liable to defensive harm; in World W_2, Bill did order the attack, and so is fully liable to defensive harm.

	Injustice: The degree to which A_1 infringes Bill's rights	Imprudence: The degree to which A_1 allows Alice to be harmed	Degree of overall vice
World W_1 ($P(W_1\|A_1) = 0.1$)	10	0	10
World W_2 ($P(W_2\|A_1) = 0.9$)	0	0	0

Act 2: Alice does not kill Bill

In World W_3, Alice does not kill Bill, and dies in the lethal attack.

	Injustice: The degree to which A_2 infringes Bill's rights	Imprudence: The degree to which A_2 allows Alice to be harmed	Degree of overall vice
World W_3 ($P(W_3\|A_2) = 1$)	0	10	10

Here, the conditional probability of W_1, conditional on Act 1, is 0.1 ($P(W_1|A_1) = 0.1$), while the conditional probability of W_2, conditional on A_1, is 0.9 ($P(W_2|A_1) = 0.9$). So, Act 1's expected degree of overall vice is 1 ($= 10 \times 0.1 + 0 \times 0.9$). Since the conditional probability of W_3, conditional on Act 2, is 1 ($P(W_3|A_2) = 1$), Act 2's expected degree of overall vice is 10. Thus, in Case 1, the act with the lowest expected degree of overall vice is Act 1 (since $1 < 10$). Act 1 therefore has the highest expected degree of overall virtue: so, Alice has most reason ATC to do Act 1.

To determine which acts are morally wrong, we need to consider the reasonable weighting of the virtues that assigns the greatest weight to the non-moral virtue of prudence, and the lowest weight to the moral virtue of justice. We may assume that the numbers given in the tables for Acts 1 and 2 above correspond to this minimally-moral weighting of the virtues. On this weighting of the virtues, the only maximally virtuous act is Act 1; and Act 2 is not worse than Act 1 in terms of the moral virtues. So, in this case, neither Act 1 nor Act 2 is subjectively morally wrong; both acts are subjectively morally permissible.

In this case, if the actual world is W_1, Act 1 is objectively wrong, but subjectively right, and it is plausible that Act 1 violates Bill's rights, although it does so excusably – Alice is not blameworthy for violating Bill's rights in this way. If the actual world is W_2, it is plausible that Bill "forfeits" his right against defensive harm, and so in this world Act 1 is neither objectively nor subjectively wrong.

Case 2: Alice is a *saintly hero*

Case 2 is just like Case 1, except that now justice is weighted more heavily, while prudence is weighted less heavily. For example, perhaps in this case, Act 1's expected degree of overall vice is 9, while Act 2's expected degree of overall vice is 8. Thus, in Case 2, the act with the lowest expected overall degree of vice – and so the highest expected overall degree of virtue – is Act 2 (since $8 < 9$). Thus, in Case 2, Alice has most reason ATC to do Act 2.

Even in Case 2, however, Act 1 is not subjectively wrong, because there is a reasonable weighting of the virtues (specifically, the weighting of Case 1) on which Alice would have *no less reason* ATC to do Act 1 than any alternative. Thus, in Case 2 just as in Case 1, Act 1 would still be subjectively permissible. Even if Act 1 is objectively wrong (because Bill has not in fact forfeited his rights against defensive harm by ordering the attack), Act 1 will therefore be

excusable and not blameworthy.

Portmore's "Questionable Man"

Here Ch(•) is an objective chance function capturing the conditional chances of the relevant worlds, conditional on Shamira's act.

Act 1: Shamira shoots Abaddon, in his present indeterminate state of mind; none of the twenty others in the marketplace is killed.

	Injustice: The degree to which A_1 infringes Abaddon's rights	Non-beneficence: The degree to which A_1 allows the twenty to be harmed	Degree of overall vice
World W_1 (Ch(W_1|A_2) = 1)	8	0	8

Act 2: Shamira does not intervene

In World W_2, Shamira does not intervene, and Abaddon detonates the bomb, killing twenty people; in World W_3, Shamira does not intervene, but Abaddon does not detonate the bomb, and so no one is killed.

	Injustice: The degree to which A_2 infringes Abaddon's rights	Non-beneficence: The degree to which A_2 allows the twenty to be harmed	Degree of overall vice
World W_2 ($\text{Ch}(W_2\|A_2) = 0.5$)	0	20	20
World W_3 ($\text{Ch}(W_3\|A_2) = 0.5$)	0	0	0

Since the conditional chance of W_1, conditional on Act 1, is 1 (*i.e.*, $\text{Ch}(W_1|A_1) = 1$), Act 1's chance-expected degree of overall vice is 8. Since the conditional chances of W_2 and W_3, conditional on Act 2, are both 0.5 (*i.e.*, $\text{Ch}(W_2|A_2) = \text{Ch}(W_3|A_2) = 0.5$), Act 2's chance-expected degree of overall vice is 10. Thus, the act with the lowest chance-expected degree of overall vice – and so the highest chance-expected degree of overall virtue – is Act 1 (since $8 < 10$). Thus, Shamira has most reason ATC to do Act 1. If we assume that all reasonable weightings of the virtues agree on this case, we may infer that Act 1 is objectively permissible and Act 2 is objectively wrong. Because of this, we should not say that Act 1 "violates" Abaddon's rights, but only that it "infringes" them. (Since Abbadon's state of mind is indeterminate in world W_1, it seems questionable whether he has fully "forfeited" his right against defensive harm – although perhaps these rights have been in a way *weakened* or *attenuated*.)

In this analysis, the indeterminacy in this case does not arise with respect to Act 1. If Shamira performs Act 1, there is only one relevant world – the world W_1 in which she kills Abbadon in his current indeterminate state of mind. In this world, Abbadon has an intermediate degree of liability to defensive harm, and so Act 1 appears at least to infringe his rights to some degree. In this analysis, the indeterminacy arises with respect to Act 2, since there are two very

different worlds in which Shamira performs Act 2 – W_2, the world in which Abbadon detonates the bomb, and W_3, the world in which he does not. In W_2, Shamira fails to a very significant degree to display the virtue of beneficence, while in W_3, there is no such failure of beneficence at all. Since there is no fact of the matter about which of these worlds would have obtained had Abbadon not been killed, these two very different degrees of beneficence must be weighted by the conditional chances of these worlds, conditional on Shamira's performing the relevant act.

Ka'eo at the dartboard

Here Ch(•) is again an objective chance function, capturing the conditional chances of the relevant worlds, conditional on Ka'eo's act.

Act 1: Ka'eo throws the dart, aiming to hit the bullseye

In world W_1, Ka'eo succeeds in hitting the bullseye, and 10,000 people receive a small reward, while the innocent person is unharmed; in world W_2, Ka'eo fails to hit the bullseye, provoking the demon to kill the innocent person, and not to give the 10,000 people their reward.

	Injustice: The degree to which A_1 infringes the innocent person's rights	Non-beneficence: The degree to which A_1 fails to help the 10,000	Degree of overall vice
World W_1 ($Ch(W_1\|A_1) = 0.5$)	0	0	0
World W_2 ($Ch(W_2\|A_1 = 0.5$)	8	2	10

Act 2: **Ka'eo stops playing darts**

In World W_3, Ka'eo stops playing darts, and the demon neither kills the innocent person nor gives the 10,000 their small reward.

	Injustice: The degree to which A_2 infringes the innocent person's rights	Non-beneficence: The degree to which A_2 fails to help the 10,000	Degree of overall vice
World W_3 ($Ch(W_3\|A_2 = 1$)	0	2	2

Since the conditional chances of W_1 and W_2, conditional on Act 1, are both 0.5 ($Ch(W_1|A_1) = Ch(W_2|A_1) = 0.5$), Act 1's chance-expected degree of overall vice is 5. Since the conditional chance of W_3, conditional on Act 2, is 1 ($Ch(W_3|A_2) = 1$), Act 2's chance-expected degree of overall vice is 2. Thus, the act with the lowest chance-expected degree of overall vice – and so

the highest chance-expected degree of overall virtue – is Act 2 (since 2 < 5). Thus, Ka'eo has most reason ATC to do Act 2. Assuming that all reasonable weightings of the virtues agree on this case, Act 2 is objectively permissible, and Act 1 is objectively wrong.

Since Act 1 is objectively wrong, it seems plausible that in this case, Act 1 would "violate" the innocent person's rights. Indeed, I am inclined to think that Act 1 would violate the innocent person's rights even in World W_1 – in which Ka'eo succeeds in hitting the bullseye, with the result that the innocent person is not killed by the demon; even in this world, Ka'eo wrongly exposed the innocent person to an unacceptable risk of being killed, which seems to be enough to violate the innocent person's rights.

In this way, an act-focused version of virtue ethics, which appeals to the virtues (such as justice, beneficence, and prudence) that are instantiated by the acts themselves (relative to the relevant possible worlds), can be married with decision theory to yield a form of DTVE. This form of DTVE seems capable of giving an adequate treatment to the problem cases that we considered in Section 1 above.

6. Conclusion

This concludes my presentation of DTVE. My goal here has simply been to present this theory, not to compare it to every possible alternative theory, or to attempt to show that it is more plausible than each of these alternatives. Since (to my knowledge) no one else has attempted to marry virtue ethics with decision theory, in the way that I have tried to do here, the first task is just to articulate the theory; it is only after the theory has been articulated that it can be evaluated and compared with all its rivals.

Still, in presenting the theory, I have tried to bring out its theoretical resources. The

theory can be developed in many different directions: for example, we could assume different lists of virtues as those that are relevant to permissibility, or different accounts of how these virtues are aggregated to yield judgments about how much reason ATC there is for each of the available acts. Even if an additive account of the aggregation of different virtues is assumed, different accounts can be given of which weightings of the different virtues are reasonable. Given how few non-consequentialist accounts have been developed of how we can be guided by probabilities in making ethical decisions, new theories for us to discuss and evaluate are urgently needed. The fact that DTVE also has these theoretical resources is a point in its favour.

As I explained above, this sort of DTVE has drawn on a number of ideas that were explored by Joseph Raz – above all, his idea of a values-based but not-necessarily-consequentialist view of reasons for action, but also some of his ideas about how some values can be "small" compared to others, and about the significance of incommensurability. While he himself never explored whether his insights could be combined with decision theory, it is in my view one of the great merits of his insights that they make room for this combination.