

Comments on Ralph Wedgwood, "Decision-Theoretic Virtue Ethics"
Pacific APA 2023
Andy Egan, Rutgers University
andyegan@rutgers.edu

1. Saying it back

Here's my understanding of how the framework is working, and what the crucial moving parts are:

The big picture is that Ralph is delivering a framework that'll let us use the tools of decision theory to get a grip on decision making under uncertainty even if we don't just care about consequences.

One important moving part:

- Actions are getting evaluated with respect to various distinct virtues
- then how well or badly the action does with respect to the various virtues is getting agglomerated into an ATC virtue-based value of the action, based on an assignment of weights to the different virtues.

So what's happening in the framework is this:

- We first do something like the standard decision-theoretic thing on the action, but we do it a bunch of times - once for each distinct virtue.
- For each virtue, we run the action through a process that's sensitive to both a that-virtue-specific value function and a probability function (the same probability function for all the virtues) to get an expected value with respect to that virtue.
- Then we take the resulting virtue-specific values and we agglomerate them into an ATC value by adding them together, where each virtue's contribution to the sum is multiplied by a factor that captures that virtue's weight relative to the others in its importance to all-things-considered choiceworthiness.
 - (And also corrects for any funny business in the units in the original assignments of virtue-specific values.)
- That gets us ATC relative values of actions, and gets us facts about which action(s) are ATC most choiceworthy given a weighting of values and an assignment of probabilities.

If we thought there was in every decision situation a single uniquely correct weighting of the virtues, I think maybe we could stop there. But there's one more step from here to moral wrongness and permissibility:

So another important moving part:

- There's an extra step to get to verdicts on wrongness & permissibility:
 - an action's being wrong is a matter of how it compares to other actions *across a range of different weightings of the virtues*.
 - So there's no single weighting of virtues that's doing all the theoretical work - facts about how the actions compare across a number of different weightings makes a difference to what the framework says about wrongness and permissibility.
- An act is morally wrong iff it does worse wrt moral virtues than every act that maximizes ATC virtue on any reasonable weighting of the virtues.
- (So iff it does worse wrt moral virtues than every act that does best wrt ATC choice worthiness, even on reasonable weightings of the virtues that give the most importance to the non-moral virtues relative to the moral ones.)

2. Questions

First question:

- Can you say some more about the motivations for handling moral wrongness in the way you do it?
- (Also: Is the story about other sorts of evaluation that aren't obviously virtue-specific going to get the same treatment? For example rationality/irrationality, prudence/imprudence? Or is moral evaluation distinctive in this way?)

Second question:

- The assumption that the contributions of the different virtues to overall choiceworthiness can be happily modeled in terms of multiplicative weighting factors seems substantive and potentially controversial. There are some views about the relative importance of different virtues that seem like maybe we should take them seriously, but can't be modeled this way.
- I'm thinking about lexical value cases - where there's some pair of virtues such that the first shouldn't be traded off against the second at *any* rate of exchange.
- (The judge who's more eloquent when announcing guilty verdicts.)

Third question:

- How virtue-ethics specific is the framework you're offering? (This is a "maybe you've done something even cooler and more broadly applicable than you've let on" question.)
- It looks like the general framework you've given is one we could use any time you've got different dimensions of value that can pull in different ways in a decision situation, and where the relative weights of the different dimensions of value is something we don't want to represent as just more facts about the world.
- Are you happy for that to be true, or do you think there's something distinctive about the virtue ethical case that makes the framework applicable here but not elsewhere?

Fourth question:

- Do you get an EDT/CDT fight about how to extract the expected virtue-specific values that feed in to the aggregation process from a virtue-specific value function and a probability function?
- I think maybe you do, and that maybe there are news-management complaints about the way you set it up in the paper.
- I think the conditions for getting an EDT/CDT fight about virtue-specific expected values going are satisfied:
 - there are going to be some background conditions that are relevant to how well an action does wrt some virtue
 - and the presence or absence of those background conditions might be evidence about what I'll do (and so what I do will be evidence about which virtue-relevant background conditions obtain).
- Cases:
 - Tyler poking me in the eye
 - Addiction & compulsion
- Does that seem right? Does it seem worrying? If the CDT/EDT debate is going to happen here too, does it all go in just the same way, or is there distinctive stuff that might make it go differently here than in standard decision theory?