CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Why Impossible Options Are Better: Consequentializing Dilemmas

Brian Talbot (ID)

Email: btalbot@colorado.edu

Abstract

To consequentialize a deontological moral theory is to give a theory which issues the same moral verdicts, but explains those verdicts in terms of maximizing or satisficing value. There are many motivations for consequentializing: to reconcile plausible ideas behind deontology with plausible ideas behind consequentialism, to help us better understand deontological theories, or to extend deontological theories beyond what intuitions alone tell us. It has proven difficult to consequentialize theories that allow for moral dilemmas or that deny that "ought" implies "can." This article argues that the problem is best solved by allowing impossible actions as inputs into consequentializations. It shows that all other approaches that have been advocated are inadequate. It also argues that progress in consequentialization, and in formal ethics more generally, requires thinking about more than just wrongness and permissibility; we should think about contrary-to-duty obligations and degrees of wrongness as well.

A deontological moral theory can be *consequentialized* iff, roughly speaking, we can give a theory that issues the same moral prescriptions as it does, but which explains these prescriptions purely in terms of maximizing (or satisficing) value. It has proven difficult to consequentialize theories that allow for *moral dilemmas*: cases in which all the actions an agent can take are impermissible (Peterson 2010, Brown 2011, Portmore 2011, Dietrich & List 2017). I will argue that all approaches to consequentializing dilemma-allowing theories that others have advocated are inadequate and that there is a superior alternative. Along the way, I will discuss features such as contrary-to-duty obligations and degrees of wrongness that have been largely overlooked in the literature on consequentializing, but which are core parts of many deontological theories. These will be of interest to anyone thinking about formal models of deontological theories, even if they are not interested in consequentialization or deny that there are dilemmas.

Prototypical moral dilemmas involve conflicts between two duties, neither of which is weightier than the other. This is what we see, for example, when Sophie is forced to choose which of her children to sacrifice (Styron 1979). But these are not the only types of dilemmas. Any counterexample to the principle that "ought" implies "can" is a moral dilemma. That is because any such example involves an obligation that

cannot be fulfilled, and thus is a situation in which all possible actions are wrong.[1] This is noteworthy for two reasons. First, as we will see, this latter sort of dilemma poses more serious issues for consequentialization than do prototypical dilemmas. Second, the evidence against "ought" implies "can" is clearer than the evidence for prototypical dilemmas. There's a long-standing line of research on linguistic data and moral intuitions that run contrary to "ought" implies "can"; the data suggests that most non-philosophers deny that "ought" universally implies "can" (this line of research runs from Sinnott-Armstrong 1984 to Buckwalter & Turri 2015 and Chituc et al. 2016). Many philosophers deny "ought" implies "can" without (clearly) accepting the possibility of prototypical dilemmas (e.g. Ryan 2003, Graham 2011, Talbot 2016). And many philosophers endorse weakenings of "ought" implies "can" that allow for non-prototypical dilemmas. One such weakening is that "ought" implies "normally can," which allows agents to be in dilemmas when they are faced with an obligation that normally, but not currently, can be fulfilled (Alston 1988, Ryan 2003).[2] Another is "ought" implies "could have": if an agent is obligated now to do such and such, then at some point in the past they could have acted to make such and such possible now (see, arguably, Stocker 1971, or Howard-Snyder 2006). This allows for dilemmas when an agent has previously made themselves unable to fulfill an obligation they currently have. Because these theories are more widely plausible, consequentializing theories that deny unrestricted "ought" implies "can" is more pressing than consequentializing theories that just allow prototypical dilemmas. It is also more difficult.

Consequentializations of theories have two necessary features. They must assign values to various options, and they must use those values to generate *deontic outputs* about acts – for example, that acts are wrong or permissible (the term "deontic output" is borrowed from Brown 2011).[3] Let's call the function that generates deontic outputs the *deontic output function*. The literature on consequentializing focuses mostly on *maximizing* deontic output functions, which (roughly) say that an act in a situation is wrong if is not the highest valued act relevant to that situation, and permissible if it is. I also focus on maximization, although the issues I discuss are relevant to "satisficing" consequentializations as well (I'll say more about this in section 2).

A successful consequentialization of theory T must at least generate the same deontic outputs as T: it must agree with T about what is wrong, permissible and so forth. There

---

[1]While the literature on consequentialization focuses on prototypical dilemmas (except Dietrich & List 2017), what I say here is entailed by standard definitions of "moral dilemma."

Some authors distinguish between prohibition dilemmas, in which all possible options are wrong, and obligation dilemmas, in which an agent is faced with multiple incompatible requirements (e.g. Vallentyne 1989). There are ways of understanding the relationship between requirements and prohibitions that collapse this distinction. They are also ways of understanding this relationship that allow this distinction and make obligation dilemmas conceptually impossible (e.g. Vallentyne 1989, Portmore 2007). I'll remain neutral on this and just focus on prohibition dilemmas.

[2]Alston's paper is about epistemic obligations, but it seems clear that he takes this principle to be plausible in moral contexts as well.

[3]Standard consequentialism assigns values to outcomes. Those who think that consequentializations must do so as well can read "assign values to *options*" as referring to evaluating outcomes. To deontologists, the ends do not always justify the means; in a sense, acts themselves can be seen as having values. To reconcile this with consequentializations that only evaluate outcomes, we include as an outcome of an act that the act was performed. For some deontological theories and some acts, no other feature of the world other than that the act is performed is relevant to the value of the outcome. None of this will affect the arguments I make.

is disagreement about additional success conditions for consequentializations, however, because there is disagreement about why consequentialization matters (see Schroeder 2017 for an overview). One motive for consequentializing is to *reconcile* consequentialism and deontology (Oddie & Milne 1991, Louise 2004, Portmore 2007). The thought is that some of the general ideas underlying consequentialism seem plausible, but much of what deontology says seems plausible as well; if we can consequentialize a deontological theory, then we reconcile these different plausible ideas. Another motive for consequentializing is to allow the use of formal tools to give us a better *understanding* of particular deontological theories (Louise 2004, Peterson 2010, Dreier 2011). Finally, we might see consequentializations as enabling *extensions* of our moral theories: our intuitions only tell us so much, and if we can formalize a deontological theory, we can determine what is wrong or permissible in cases that go beyond our intuitions. For example, consequentializing deontological theories allows us to apply decision-theoretic tools to tell us what agents should do when they are uncertain, which is otherwise a vexing issue for deontology (Peterson 2010, Colyvan et al. 2010, Lazar 2018). My arguments below are agnostic between these motivations; I discuss the different success conditions on consequentializations that we get given each. I also treat a wider range of deontic outputs than usually discussed. Those writing on consequentializations have focused just on wrongness and permissibility, but as we will see, deontological theories are often about more than just these. No matter which motivation we have for consequentializing, a successful consequentialization of a theory must be able to account for all that theory's different kinds of deontic outputs.

Here's how the article will go. Section 1 explains why moral dilemmas initially seemed like a problem for consequentialization, and then shows how this problem has been solved. It then articulates a new problem based on non-prototypical dilemmas and a kind of deontic output not discussed in the consequentialization literature – contrary-to-duty obligations. Section 2 talks about three approaches to consequentialization that seemingly solve this new problem. Section 3 argues that only one really solves it, and section 4 buttresses that by discussing an additional kind of deontic output – degrees of wrongness – and how ideas from philosophy of science can be brought to bear on consequentialization.

1

Let's start with the initial problem that moral dilemmas pose for consequentializations. If an agent has a choice between multiple options, it seems like at least one of them must maximize value. So, according to consequentialism, it seems like an agent always has at least one permissible option. But in a dilemma, all acts the relevant agent can perform are wrong. How can a consequentialization capture this?

The initial response to this problem is what I will call the *multi-ranking approach*. According to the multi-ranking approach, in a dilemma there are multiple, different valuations of the options, each of which assigns a different option the highest rank. Further, the deontic output function says that permissibility requires being on top of all relevant rankings, so in these cases no option is permissible.

One type of multi-ranking consequentialization is discussed by Campbell Brown (2011) and endorsed by Martin Peterson (2010). It appeals to the plausible idea that moral dilemmas can be generated when we have incompatible and incomparable obligations. To model this, it says that rankings can be incomplete – they can leave out options. When an agent choosing between A and B is in a dilemma, there are two

rankings. One puts A on top, but does not rank B at all. The other puts B on top, and does not rank A. Neither A nor B is ranked highest on all relevant rankings, so neither is permissible. An alternative type of multi-ranking consequentialization is given by Doug Portmore (2011). It says that how options are valued can depend on what option an agent chooses. In a dilemma where A and B are the options, there is a ranking of options given that the agent chooses A, and a ranking given that the agent chooses B; the ranking given the choice of A puts B first, and the ranking given the choice of B puts A first. No option is maximal according to all relevant rankings, so the agent does something wrong no matter what.[4]

Multi-ranking approaches can issue the proper outputs about wrongness and permissibility in dilemmas.[5] But they run into difficulties with another important sort of deontic output: *contrary-to-duty obligations*.[6] An agent in a dilemma has a contrary-to-duty obligation when they are required to take one of the available options over another, even though all options are wrong. To be clear, there's a sense in which agents in dilemmas are always required to do something wrong. This is the sense that we get if we think of "*x* is required" as meaning "it is wrong to not do *x*." All options in a dilemma are required in this sense. Contrary-to-duty obligations in dilemmas go beyond this. They occur when it is clearly morally preferable (speaking loosely) to take one of the wrongful options in the dilemma over the others.

We can illustrate this with some examples. Imagine that Agnes is a kleptomaniac who cannot help but steal. However, when she steals, she is capable of choosing to steal less valuable items rather than more valuable items.[7] Some plausible moral theories will say that it is wrong for Agnes to steal, even though she can't help herself (e.g. Ryan 2003, Graham 2011, Talbot 2016). This will sometimes put her in a dilemma, since sometimes all her possible choices will be wrong to take. Given that Agnes is going to steal at some time, she is required to steal a less valuable item rather than a more valuable item. Although both are wrong, there's something different, and more important, about the option of stealing the less valuable item; Agnes has a contrary-to-duty obligation to take this option. Some may want to deny this on the grounds that Agnes' incapacity is not her fault. Consider instead Bahi, who is a security guard in a museum. For no good reason, she starts swinging her heavy flashlight negligently near some art. She realizes that she has lost control of the flashlight and is going

---

[4]Jussi Suikkanen (2019) discusses an additional form of multi-ranking from Portmore (2011). This says that options can be ranked both by their moral value and by their all-things-considered value. Portmore does not use this to model dilemmas, but Suikkanen shows that it can be so used if we say that options are wrong if they are not ranked highest on both lists. This will have the same problems as the view of Portmore's that I state in the main text.

[5]Dietrich & List (2017) argue that multi-ranking approaches cannot handle all dilemmas. There can be cases where an agent has only a single available action, and this action violates some impossible-to-fulfill obligation. Multi-ranking approaches cannot use multiple, inconsistent rankings to account for these dilemmas. In response, defenders of multi-ranking approaches might adopt a very fine-grained notion of "options" which says that agents always have multiple options. For example, if one has no choice but to miss an appointment and break a promise, one might miss the appointment while chewing gum or miss the appointment while not chewing gum, which count as different options and get ranked differently. This might allow multi-ranking approaches to generate the correct verdicts about wrongness and permissibility in all dilemmas.

[6]See Frances Howard-Snyder (2006) for an argument that, if we accept dilemmas, we need to also explain contrary-to-duty obligations. She thinks that we cannot do this. My discussion in sections 2–4 shows why I disagree.

[7]This is a modified version of Forrester's (1984) example of gentle murder.

to smash a piece of art; all she can do is slightly alter the course of her swing to smash either a more valuable or less valuable piece. She has a contrary-to-duty obligation to smash the less valuable piece, but smashing this piece is still wrong. Note that we get this result even if we accept a weak version of "ought" implies "can," such as "ought" implies "normally can" or "ought" implies "could have." Agnes' and Bahi's cases each involve a single type of duty – the duty to not steal or not negligently break art – which can vary in strength. We can also get contrary-to-duty obligations in dilemmas involving more than one type of duty. Imagine that Carlos is babysitting a toddler. He intentionally feeds the child deadly pills. Then, out of a desire to avoid punishment, he decides to save her life. The only way to do so is to steal a neighbor's car to drive to the hospital. This theft is intuitively required and morally preferable to Carlos' other option of doing nothing. But the theft is wrong because the situation is due to Carlos' initial wrongdoing (doing nothing is of course also wrong). Carlos has a contrary-to-duty obligation to steal the car.[8]

Contrary-to-duty obligations are not an issue if we focus on prototypical dilemmas involving conflicting duties which are equally strong or incomparable. Given the equal strength or incomparability of these duties, there is no second-best option that one is required to take. But contrary-to-duty obligations are inevitable if an agent has an obligation they cannot fulfill, but where they can mitigate the wrong done, or if the agent is at fault for creating a conflict involving duties of clearly different strengths.

Why are contrary-to-duty obligations a problem for the multi-ranking approach? To explain why it is wrong for Carlos to steal the car, Peterson's approach will employ multiple but incomplete rankings of Carlos' options. One ranking will include car theft but not doing nothing, and the other will include doing nothing but not car theft. Both options will be wrong because neither is optimal according to all relevant rankings. While this gets the wrongness judgement correct, it cannot generate the deontic output that Carlos is obligated to steal the car. After all, it never ranks this option ahead of doing nothing. Portmore's approach, on the other hand, will appeal to multiple rankings which disagree about which action is optimal: if Carlos steals the car, doing nothing is ranked higher, and if he doesn't, then stealing the car is ranked higher. This does not explain Carlos' contrary-to-duty obligation. If the obligation to steal were explained by the fact that it would be ranked higher were Carlos to do nothing, then we would also have to falsely say that Carlos has an obligation to do nothing, as this would be ranked higher were he to steal the car. More generally, contrary-to-duty obligations in dilemmas are a problem for the multi-ranking approach because, since the cases involve dilemmas, all the possible options are wrong and so must fare badly in some comparison. But, since there is a contrary-to-duty obligation, at least one option must fare well relative to the other options. This last feature is the problem, as multi-ranking approaches have to use multiple rankings, which make different and incompatible comparisons, to account for the dilemma.

The failure to account for contrary-to-duty obligations in dilemmas is a bad result no matter what motivation we have for wanting to consequentialize. It means that we can reconcile the plausible ideas behind consequentialism with the ideas behind these forms of deontology only by giving up some of the deontic outputs that make these deontological theories plausible. For understanding and extension-based motivations, it means that we cannot use consequentializations to fully understand certain deontological

---

[8]My thanks to Julia Staffel for help with this example.

theories or accurately extend what they will say, because we cannot use consequentializations to understand or extend an important class of deontic outputs.

2

There are ways to consequentialize moral theories that allow dilemmas and contrary-to-duty obligations. One way is to modify multi-ranking approaches so they can better deal with contrary-to-duty obligations. Another is to adopt an approach suggested by Franz Dietrich and Christian List. I will spend the rest of the article arguing that both of these are inferior to a third way.

To fix multi-ranking approaches, we add another ranking.[9] We have the multiple rankings of options that we use to explain dilemmas and an additional, separate ranking which we use to explain contrary-to-duty obligations. To illustrate, consider what a modified version of Peterson's view would say about the kleptomaniac who can choose between stealing something of greater or lesser value. To make this a dilemma, there are two rankings which explain the wrongness and permissibility of her actions. One only ranks stealing the greater-valued item, and the other only ranks stealing the lesser-valued item. Since neither option is at the top of both rankings, neither is permissible. There's also a third ranking, relevant only to contrary-to-duty obligations, which ranks both options and puts stealing the lesser-valued item on top; this means that Agnes is obligated to steal this item rather than the item with greater value, but the theft is not permissible.

Alternately, we can use an approach suggested by Franz Dietrich and Christian List (2017). On their approach, some options have *non-reflexive value*. An option has non-reflexive value iff it is not at least as good as itself. The approach defines maximization of value so that an option maximizes, and is permissible, only if it is at least as good as every available option, including itself. Options with non-reflexive value thus never maximize. The view accommodates dilemmas by saying that in them all the options, or perhaps just the best option, have non-reflexive value, so no option maximizes and thus none is permissible. It can still say that agents are obligated to take a highest valued possible option, which gives us contrary-to-duty obligations.

My preferred solution says that sometimes some impossible options have values. For example, in Agnes' case we could assign value not only to stealing lesser and greater-valued items, but also to the impossible (for Agnes) act of not stealing. Further, the approach says these values are sometimes relevant to deontic outputs – that is, sometimes the deontic output function takes into consideration the value of some impossible options. When impossible options are relevant to deontic outputs, and are valued more highly than all possible options, there will be a moral dilemma, since no possible option maximizes and thus no possible option is permissible. The highest ranked possible option will be obligatory but still wrong, so the view delivers contrary-to-duty obligations. What do I mean by "impossible?" The details depend on the theory to be consequentialized. We get dilemmas when there is nothing an agent can do that is permissible. Different theories will have different views about what that "can" means. By "impossible," I mean "cannot be done" in that sense.

Let's briefly consider two more approaches to consequentialization that can deal with dilemmas and contrary-to-duty obligations. To work, these must incorporate impossible options, and so they are really versions of the impossible-options approach.

---

[9]My thanks to an anonymous referee for urging me to discuss this.

Jussi Suikkanen ([2019](#)) discusses using satisficing consequentialism to model dilemma-allowing theories. Satisficing consequentialism takes one of two forms. One form uses value thresholds to determine wrongness and permissibility. To model dilemmas, we say that any option below a certain value threshold is wrong; dilemmas arise when all options are below this threshold. Threshold-based approaches to consequentialism face a variety of devastating objections. Suikkanen shows how they over-generate dilemmas. They also over-generate permissions: for example, if one, easily taken, option allows a huge amount of value to be generated, way over the threshold, but another option prevents that from happening while still remaining slightly above the threshold, the view will implausibly say both are permissible (Bradley [2006](#)). Further, the understanding and reconciliation motivations for consequentializing need consequentializations to involve comparisons between options (as I discuss in the next section). So, threshold-based consequentializations should be rejected. The second form of satisficing consequentialism determines the wrongness and permissibility of an option by comparing it to one's other options. Suikkanen says that this will also generate too many dilemmas. But that's because he considers a version according to which an option is permissible only when it is much better than all other options. That is a bad view. But there's another way to use satisficing consequentialism to generate dilemmas: allow impossible options, and say an option is wrong when it has much less value than another option. We then get dilemmas when all possible options have significantly less value than some impossible option. So, satisficing consequentialism can model dilemmas as long as it is combined with the impossible-options approach.

Some philosophers use dyadic versions of deontic logic to deal with contrary-to-duty obligations (see Lewis [1974](#), Prakken & Sergot [1997](#)). This approach sees obligations as relative to contexts: one is obligated (in a context) to do what is best in that context. Obligations are relative to "ideal" contexts, and contrary-to-duty obligations are relative to non-ideal contexts (i.e. contexts where we don't fulfill our obligations). We can use this to model dilemmas. But to capture certain dilemmas, we have to incorporate impossible options. Think about Agnes, who cannot help but steal. There's no possible context in which she cannot do the best possible thing. If the options and contexts that determine obligations were always possible, then there would be a possible option that satisfied her obligations, and so she wouldn't be in a dilemma. To capture cases like hers, we have to judge her actions relative to impossible options. More generally, to model any dilemmas that violate both "ought" implies "can" and "ought" implies "could have" using dyadic deontic logic, we have to incorporate impossible options.

So, we have three general approaches to consequentializing theories that allow dilemmas: multi-ranking approaches, non-reflexive value approaches, and impossible-options approaches. Each general approach has variants – e.g. Peterson and Portmore's versions of multi-ranking, or maximizing, satisficing, or dyadic versions of the impossible-options approach. My description of these general approaches has been quite abstract. That is partly because they should all be able to be reconciled with a wide range of deontological theories, each of which will require details to be specified in different ways. Another reason I've left things so abstract is that the particulars of these approaches will depend on how we address challenges faced by any view that allows dilemmas. The rest of this section will discuss these challenges, because they represent fundamental worries about consequentializing dilemma-allowing theories.

Not every value can be non-reflexive, nor can every decision involve multiple conflicting rankings, nor can the value of every impossible action be relevant to wrongness in every situation. Any of these would generate too many dilemmas. So, all three

approaches need an account of when the dilemma-generating formal feature is applied to a given choice. The first question that arises is: can any such account be given at all? I will assume so. This is an instance of a more general issue for moral theories that deny that "ought" implies "can": any such theory needs an account of when we are and are not obligated to do the impossible (Howard-Snyder 2006, Graham 2011, Talbot 2016, Suikkanen 2019). Interest in consequentializing dilemma-allowing theories assumes that such an account can be given. We don't have any such account at the moment (although see, for example, Graham 2011 and Talbot 2016 for tentative suggestions), but there is not much evidence that no account is possible, given that so few have tried to find one. The second question that arises is: will any of these three approaches be a better fit to such an account, whatever it is? I don't see any reason to think that one is better than the other in this respect. The only necessary difference between the approaches is *whether* they use multiple rankings, *whether* they assign any options non-reflexive values, and *whether* impossible options are ever input into the deontic output function. Those three features obviously cannot be used to explain *when* multiple rankings are used, *when* values are non-reflexive, or *when* impossible options are ranked. So, any account that can be given to say when dilemmas arise will have appeal to other features of the consequentialized theory, which need not differ between consequentializations that use the non-reflexive-value approach, those that use the multi-ranking approach, or those that use the impossible-options approach to dilemmas.

This points to another problem for consequentializing dilemma-allowing theories, however. Let's say we can give a successful account of when to apply the dilemma-generating formal feature. Suikkanen (2019) argues that our very ability to do so undermines the goals of consequentializing. That's because giving this account will rely on our prior understanding of what generates dilemmas – that is, understanding we have prior to giving the consequentialization – and thus the consequentialization will not help us to better understand this aspect of our deontological theory.[10]

I have a number of things to say in response. For one, I suspect that consequentializing *will* help us understand when and why dilemmas arise. If philosophers eventually produce plausible accounts of when we have impossible-to-fulfill obligations, I suspect that these will initially be somewhat imprecise. That's just how philosophy tends to go. Trying to consequentialize theories that include these accounts will help us see how they are imprecise and how to precisify them; that will give us more understanding of dilemmas.[11] But perhaps I'm mistaken, and consequentialization will not help us understand when we are obligated to do the impossible. That's sub-optimal, but it does not negate the goal of consequentializing. When we consequentialize a theory to better understand it, we do not expect the consequentialization to illuminate every aspect of that theory. For example, when we consequentialize theories that do *not*

---

[10]Suikkanen has another concern: we also won't be able to use our consequentialization to better justify the initial deontological theory, because it will be based on the same data as the deontological theory. I'm not convinced. If we can use the consequentialization to reconcile the plausible ideas behind consequentialism and deontology, this seems to provide further support for the view.

[11]We see parallel things in other areas of philosophy that have adopted formal methods. Take epistemology, for example. The idea that epistemic norms, such as the norm of coherence, derive in some way from the pursuit of truth is historically very important, but relatively imprecise. There has been a great deal of work in formal epistemology to make this idea and these norms more precise. This discussion starts with our pre-theoretic understanding of epistemic goals, values, and norms, builds formal models around them, and ends up giving us a much more precise understanding of things like accuracy and coherence, of exactly how norms should work, and why (see e.g. Pettigrew 2016).

allow dilemmas, we employ a deontic output function that only ranks possible options. Why? Because (roughly) these theories accept that "ought" implies "can." That acceptance is prior to the consequentialization, and the consequentialization sheds little light on it. Even so, those who advocate consequentializing with the goal of understanding deontological theories don't think that there is no point to consequentializing theories that do *not* allow dilemmas; consequentializing helps us understand other parts of the theory. So there can still be a point to consequentializing theories that *do* allow dilemmas, even if doing so doesn't shed light on why dilemmas occur. Finally, Suikkanen's objection does not apply to other motivations for consequentializing. One alternate goal of consequentialization is to reconcile the plausible ideas behind consequentialism and deontology. Achieving this does not require that consequentializing a theory teach us anything about that theory. Another alternate goal of consequentializing is to help extend a theory to cases about which we have no intuitions. Even if we construct our account of when impossible options get ranked based on cases we do have intuitions about, we can still presumably apply this to cases we have no intuitions about. I will address how the different approaches to consequentialization fit with these motives below.

It looks like multi-ranking, impossible-options, and non-reflexive-value approaches can in principle issue the same deontic outputs about wrongness, permissibility and contrary-to-duty obligations in the same cases; they would just explain these in different ways. So, to this point, the views seem equally good.[12]

3

But in fact we should prefer the impossible-options approach. To see why, let's consider the motivations for consequentializing mentioned earlier in the article.

One motivation is to reconcile the plausible ideas behind consequentialism with the plausible ideas behind deontology. Given this motivation, the impossible-options approach is superior. One of the core thoughts behind consequentialism is that deontic outputs can be captured and explained by comparisons: that is, acts don't intrinsically have their deontic status, but rather are wrong or permissible given certain alternatives. A related core thought behind consequentialism is that we should maximize value (this is related because maximization is a comparative notion) (Portmore 2007).

The non-reflexive-value approach distorts both ideas beyond recognition. It says that acts in dilemmas are wrong because they compare unfavorably to themselves. This is not the kind of comparison that makes consequentialism intuitively compelling. Nor is it the kind of failure to maximize value that we have in mind when we think it is wrong to not do the best thing. The non-reflexive-value approach also seems to distort the plausible ideas behind deontology. While it issues correct verdicts, its explanation of these verdicts involves a bizarre formal contrivance – non-reflexive value – that would not be recognizable to any deontologist.

The multi-ranking approach also distorts the notion that deontic outputs have to do with maximizing value. It says that wrongness and permissibility are based on value maximization, but contrary-to-duty obligations are about maximizing something else.

---

[12]A complete consequentialization can also use some combination of these three approaches. It may be that some dilemmas really are due to incomparable duties, and would be better consequentialized using a multi-ranking approach. I'm open to that. However, if we do allow non-prototypical dilemmas, these should be consequentialized using the impossible-options approach.

That's because the latter are based on a totally different ranking than the former, and the latter ranking has no bearing on the former outputs. This thing which is not value, in the sense of "value" relevant to wrongness or permissibility, is important enough to generate contrary-to-duty obligations but not important enough to make actions wrong or permissible. The positing of such a thing is a poor fit to consequentialism.[13] Consequentialism is open to different kinds of value, and even to incomparable types of value, but no form of consequentialism talks about the moral relevance of non-value.[14]

Multi-ranking consequentialism also distorts plausible ideas behind deontology. To see what I mean, think about Bahi's case. Due to her wrongful negligence, she cannot stop herself from smashing art, and has to choose between more or less valuable pieces to smash. Contrast Bahi with Bahi*, who *could* stop herself from smashing the art, but refuses to do so. Bahi* is not in a dilemma, since she can do something permissible (not smash the art). But, given that she won't do the permissible thing, she has (like Bahi) a contrary-to-duty obligation to smash the lesser-valued art. We can think about how Bahi*'s options compare to one another, and this sort of comparison is perfectly apt for deontologists to make whether or not they consequentialize. Not smashing anything is optimal (speaking loosely), smashing the lesser-valued art is second best, and smashing the greater-valued art is worst. Importantly, deontological theories are going to say that, for Bahi*, the options to smash the greater- and lesser-valued art are straightforwardly comparable, and this comparison explains Bahi*'s contrary-to-duty obligation. The only thing that seems to change when we switch from Bahi*'s case to Bahi's is that *not* smashing anything is now no longer possible, but (intuitively) nothing changes about the comparison between smashing the greater- and lesser-valued art pieces. Multi-ranking approaches deny that. In the non-dilemma version of the case (involving Bahi*), multi-ranking approaches can explain both what is wrong and what is contrary-to-duty obligatory with a single ranking, which ranks smashing the lesser-valued art ahead of smashing the greater-valued art. But multi-ranking approaches say that this comparison between these options no longer holds in the dilemma version of the case. The removal of the option to smash nothing somehow creates three different rankings of the possible options, which do not agree about how smashing the lesser-valued art compares to smashing the greater-valued art. And removing the option to smash nothing also somehow makes the ranking which explains the contrary-to-duty obligations different from the rankings that explain what is wrong to do. That is out of line with plausible deontological ideas. So, multi-ranking approaches give up some of what seems plausible about consequentialism and some of what seems sensible about deontology.

The impossible-options approach, however, fits much better with the insights behind both consequentialism and deontology. Consequentialism is a comparative approach to ethics. While comparisons between an act and an impossible alternative are not standard, the comparisons involved in the impossible-options approach are still fairly natural:

---

[13]An anonymous reviewer notes that the kind of value that multi-ranking approaches have to appeal to in order to explain contrary-to-duty obligations looks to them more like real value, and that the types of "value" used to explain wrongness verdicts look more like non-value. If that's correct, it's another form of this problem for multi-ranking views: they distort the idea behind consequentialism by using something other than value to explain wrongness verdicts.

[14]"Global" forms of consequentialism do evaluate things other than actions – character traits, strategies, and so on (Driver 2011). But this evaluation is always done by appeal to the same stuff – value – that makes actions wrong or permissible.

it is natural to say that what Agnes really should do is not steal, or that what Carlos really should do is not be in this situation in the first place. And the notion of maximization at play is very close to the standard notion of maximization used in consequentialism. It makes use of the same, perfectly ordinary, notion of better than; it simply ranges over a wider comparison class. Now consider ideas behind deontology. Deontologists often explain why something is wrong by comparing it to other things one could have done (Kamm 2007). Further, one standard understanding of disrespect is inherently comparative – to disrespect some valuable x is to treat it as less valuable than it actually is (Darwall 1977). While this is normally understood by comparing the actual treatment to an alternative, possible, respectful treatment, if we are inclined to accept dilemmas then it makes sense to understand the disrespect of actions in dilemmas by comparing the best possible option to some other impossible kind of treatment.

So, those who accept the reconciliation motivation for consequentialization should clearly prefer the impossible-options approach. Let's turn to other motives for consequentializing. Some want to consequentialize to increase our understanding of deontological ethics, by making clear how deontological theories work. Others want to consequentialize so that we extend our theories, issuing deontic outputs in areas where our intuitions are unclear. At first glance, it seems like both goals are better achieved by views which accurately map onto the structures underlying deontic outputs: an inaccurate theory would seemingly reduce our understanding, and potentially also lead to inaccurate extensions that we could not correct (we could not correct them since they would occur in regions we have no intuitive data about). As I have just argued, the impossible-options approach better captures the underlying structure of deontological theories that allow dilemmas with contrary-to-duty obligations. So, it initially seems that those who want to use consequentialization to understand or extend dilemma-allowing theories should prefer the impossible-options approach.

4

But that may be too fast. There is a lot of work in other areas of philosophy, especially philosophy of science, on formal modeling and understanding. Some of this work addresses the connection between accurate representation of the structures underlying a phenomenon of interest and understanding of that phenomenon. Many think that these go together, all else being equal (e.g. Craver 2006). If they are correct, then please read this section of the article as pointing out a sort of bonus advantage, not related to understanding, of the impossible-options approach to consequentialization. But others have argued that certain types of inaccuracies in a theory can help us understand a phenomenon of interest better. Sometimes that's because things like idealizations or abstractions help limited agents like us to better think about or use a theory (e.g. Potochnik 2017, Rancourt 2017). I want to set aside ease of use, however: consequentialism is standardly seen as giving us a criterion of rightness, rather than a decision procedure, and for this reason I suspect that ease of use won't be a strong a reason to prefer an approach to consequentializing (nor do I see any reason to think that the impossible-options approach is harder to use than the alternatives). More relevant to my discussion is the idea that certain types of inaccuracies in a model may be necessary for certain types of explanation (e.g. Batterman 2009). On this view, the fact that a formal model can still fit the data, despite eliminating or misrepresenting details about a phenomenon, shows that these details aren't really part of the explanation of the

phenomenon in a modally robust way. That is, they aren't a necessary aspect of explanation, even if they play some contingent role in explaining actual events. On this view, inaccurate models can show us which general features fundamentally explain a phenomenon across possible worlds.[15] Inaccuracies about the underlying structure of a particular theory could thus increase understanding. Given this possibility, when we think about how consequentializations help us understand deontological theories, we cannot just think about how accurately they represent the theories. Instead, we should also think about how modally robust they are.

I will use a test case to argue that the impossible-options approach is actually more modally robust than the alternatives. The test case involves a kind of deontic output that won't be relevant to every deontological theory. That's okay, because the way to show modal robustness is to show that an approach properly models a range of deontological theories. Further, this test case gives us evidence to bolster worries about the multi-ranking and non-reflexive-value approaches that I discussed at the end of section 3.

Intuitively, a particular wrong act in one situation can be more wrong, or more seriously wrong, than a different type of wrong act in a different situation. These *degrees of wrongness* are very important for some deontological theories: they partly explain how much we are justified in punishing, the strength of our reasons to apologize or make restitution, or the strength of our reasons to avoid complicity in others' wrong actions. Not every deontological theory allows for degrees of wrongness: the sorts of deontological theories that do not tend to be Kantian or absolutist approaches that also do not allow for dilemmas (Calder 2005). But consequentializing many plausible theories does require deontic output functions that can tell us how wrong acts are. Theories which allow for degrees of wrongness will explain them in one of two ways (Hurka 2019). The *absolute-strength view* says the degree of wrongness of an act is explained solely by the strength of the duty the act violates, and doesn't involve comparisons to alternative options. The *gap view* says that the degree of wrongness of an act is a function of the difference between the wrong act done and some alternative act: for example, saving one person when you could just as easily (with a single act) have saved two is wrong, but less wrong than saving one person when you could just as easily (with a single act) have saved three.[16]

The impossible-options approach to consequentializing dilemmas can fit both the gap view and the absolute-strength view. To fit the gap view, it says that the degree of wrongness of acts in dilemmas is explained by the gap between the value of these acts and some impossible alternative. How does it fit the absolute-strength view? Acts in dilemmas violate duties whose strengths are presumably captured somehow by the values we plug into the deontic output function, which gives us all we need to accommodate the absolute-strength view.

The non-reflexive-value and multi-ranking approaches can both fit the absolute-strength view in the same way as the impossible-options approach: in each, acts in

---

[15]This is, I suspect, one of Dreier's (2011) motivations for consequentializing: consequentializing would show that uniquely deontic features of theories don't play a "real" explanatory role in ethics.

Most philosophers think that ethical truths are necessary. If ethicists are interested in modally robust explanations, it's a different, probably epistemic, sense of modality.

[16]Not much has been written about these views. This example is adapted from Graham (2016). Hurka (2019) prefers the absolute-strength view but concedes that there is something plausible about the gap view. Other philosophers I've spoken to find the gap view convincing, and it seems to map well onto common-sense morality.

dilemmas violate duties whose strength is reflected in the values of the possible options. But the non-reflexive-value and multi-ranking approaches cannot be reconciled with the gap view of degrees of wrongness.

On the non-reflexive-value approach, when someone chooses the best option in a dilemma case, what one does is wrong because one has chosen something that is not as good as itself. However, while any non-reflexive value is not as large as itself, there is not any particular amount by which it diverges from itself, nor (according to the non-reflexive value approach) is there a relevant and superior alternative to which we can compare it. So, if we combine the gap view of degrees of wrongness with the non-reflexive value approach to consequentializing dilemmas, we have to say that either there are no degrees of wrongness in dilemmas, or that it is always equally wrong to choose the best option in any dilemma. Neither is correct.

According to Peterson's version of the multi-ranking approach, we have rankings which explain wrongness and rankings which explain contrary-to-duty obligations. In the rankings which explain wrongness, the agent's options in a dilemma never appear on the same ranking, so there is no comparison between any wrong option and any other wrong option. In the rankings that generate contrary-to-duty obligations, the obligatory act is ranked best but is still wrong, so there's no gap between it and a superior option to say how wrong it is. Peterson's approach cannot fit the gap view.

Things are more complex when we consider combining Portmore's approach and the gap view. It'll be easiest to explain this by focusing on an example from earlier in the article. Carlos can either steal a car to save the life of the child he intentionally poisoned or simply let the child die. Assuming that this is a dilemma, Portmore's view says that there is a ranking according to which stealing the car is better than letting the child die and a ranking according to which letting the child die is better than stealing the car. If we adopt the gap view, the first ranking is not going to tell us how wrong stealing the car is, since it puts stealing the car above letting the child die.[17] So, to see how wrong stealing the car is, we must use the ranking that says it is worse than letting the child die. However, there is no intuitive notion of badness according to which stealing the car is worse than letting the child die; we only know that stealing must be ranked below letting the child die because intuitively this is a dilemma. Since this ranking does not tap into an independent understanding of the values of these options, there is nothing that tells us how much "worse" stealing is than letting the child die, other than our prior understanding of how wrong stealing is in this case. In other words, to determine the degree of wrongness of this act, we have to already know how wrong it is. This sort of problem will arise for every dilemma in which an agent has contrary-to-duty obligations. That's because the option that the agent has a contrary-to-duty obligation to perform will be intuitively superior to all other possible options (that's why it is obligatory), but to determine how wrong it is, we have to use a ranking that puts it below another option. That ranking will thus not use an intuitive conception of value, but rather has to be derived from our antecedent knowledge of degrees of wrongness. So, if we adopt the gap view and Portmore's approach, consequentializing would not help us understand why certain wrong options in dilemmas are as wrong as they are, as it tells us nothing about their degrees of wrongness that we don't already know, nor does it represent this knowledge in a helpful way. And consequentializing

---

[17]Further, this ranking is also the one that tells us how wrong letting the child die is. Presumably, letting the child die is more wrong than stealing the car, so we cannot use the same ranking to determine the degree of wrongness of both acts.

would not allow us to extend our knowledge of degrees of wrongness to cases our intuitions don't already tell us about. The impossible-options approach is superior, since it allows us to determine degrees of wrongness by appeal to the intuitive value that each relevant option has, which need be no different when the option is possible or not.

To be clear, the problem I point out for Portmore's approach is shared to some degree by all approaches to consequentializing. For any consequentialization, some things will need to be brute and unexplained. Evaluating how well consequentializations help us understand morality thus has to be a matter of degree. The impossible-options approach lets us derive more of the consequentialization from fewer, more general principles. Portmore's approach would require us to rely on more and more specific knowledge – knowledge of degrees of wrongness in specific cases must be appealed to, in addition to our knowledge of what is wrong. The impossible-options approach allows us to do more to explain why degrees of wrongness are what they are. They are explained by the strengths of the relevant duties, rather than by saying, "They are what they are." In section 2, I said that a consequentialization need not help us understand *everything* about the consequentialized theory. Here, I am not saying that we should reject Portmore's approach just because it does not help us understand degrees of wrongness. Rather, my criticism is that it does not help us understand degrees of wrongness *and* the impossible-options approach does (to a greater degree). If we consequentialize to better understand ethics, then all else being equal, we should pick the approach that gives us more understanding.

To summarize: we might consequentialize to increase our understanding or to extend our ability to issue deontic verdicts. Many will think that these require consequentializations that are faithful to the underlying structure of the relevant deontological theory. If you agree, then section 3 showed that the impossible-options approach is the way to go, and this section just demonstrated bonus benefits of the impossible-options approach. However, one might have wondered whether consequentializations that distort the underlying structure of deontological theory could be more modally robust than those that do not, allowing deeper understanding of morality. This section gives us good evidence that multi-ranking and non-reflexive value approaches are not more modally robust than impossible-options approaches; at least, not in a way that increases our understanding or extends our theories. To dispute this, one would have to deny that the gap view is at all a live option, and also find something that the impossible options view cannot model. If multi-ranking and non-reflexive approaches are not more modally robust than impossible-options approaches, then the ways they distort deontological theories are unmotivated and undesirable.

5

The impossible-options approach better reconciles the plausible ideas behind deontology with the plausible ideas behind consequentialism. If one thinks understanding is furthered by models that accurately reflect the underlying structure of the modeled system, then the impossible-options approach gives us more understanding of deontological theories that allow dilemmas. If one thinks that inaccurate models can further understanding, the non-reflexive-value and multi-ranking approaches are not more modally robust than the impossible-options approach. So, the ways in which they misrepresent the underlying structures of deontological theories do not lead to a deeper understanding of deontology. Nor do they help us extend our theories. And we have

evidence that the inaccuracies in these views are likely to generate false extensions which we cannot detect because they occur outside the space of the intuitive. We should prefer the impossible-options approach to consequentializing dilemmas, regardless of our motivation for wanting to consequentialize deontological theories.

The arguments for these conclusions teach us lessons we can apply outside of discussions of dilemmas, or even of consequentialization; they should inform how we think of formalized deontological ethics more generally. For example, there are a number of recent discussions of how to formally model deontological theories that allow for lexical priorities of duties – for example, theories that say no number of duties to keep promises can outweigh a single duty to not take a life (e.g. Lee-Stronach 2018, Lazar & Lee-Stronach 2019). It's not enough to give formal models which issue the correct verdicts about wrongness and permissibility. Our model must also deliver a range of other deontic outputs, such as contrary-to-duty obligations or degrees of wrongness. It must also properly explain these outputs. What it means to properly explain, and whether we think a given theory does properly explain, depends on whether we accept that modal robustness can be more important than accuracy. Expanding our horizons on which deontic outputs must be explained and how explanation works will, hopefully, allow for progress in consequentialization and formal ethics more generally.[18]

References

Alston, William P. 1988. The Deontological Conception of Epistemic Justification, *Philosophical Perspectives*, **2**, 257–99.

Batterman, Robert W. 2009. On the Explanatory Role of Mathematics in Empirical Science, *The British Journal for the Philosophy of Science*, **61**, 1–25.

Bradley, Ben. 2006. Against Satisficing Consequentialism, *Utilitas*, **18**, 97–108.

Brown, Campbell. 2011. Consequentialize This, *Ethics*, **121**, 749–71.

Buckwalter, Wesley & Turri, John. 2015. Inability and Obligation in Moral Judgment, *PLoS ONE*, **10**.

Calder, Todd. 2005. Kant and Degrees of Wrongness, *The Journal of Value Inquiry*, **39**, 229–44.

Chituc, Vladimir, Henne, Paul, Sinnott-Armstrong, Walter, & De Brigard, Felipe. 2016. Blame, Not Ability, Impacts Moral "Ought" Judgments for Impossible Actions: Toward an Empirical Refutation of "Ought" Implies "Can," *Cognition*, **150**, 20–25.

Colyvan, Mark, Cox, Damian, & Steele, Katie. 2010. Modelling the Moral Dimensions of Decisions, *Nous*, **44**, 503–29.

Craver, Carl F. 2006. When Mechanistic Models Explain, *Synthese*, **153**, 355–76.

Darwall, Stephen L. 1977. Two Kinds of Respect, *Ethics*, **88**, 36–49.

Dietrich, Franz and List, Christian. 2017. What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories, *Philosophical Review*, **126**, 421–79.

Dreier, Jamie. 2011. In Defense of Consequentializing, in *Oxford Studies in Normative Ethics*, ed. by Mark Timmons (Oxford: Oxford University Press), 97–119.

Driver, Julia. 2011. *Consequentialism* (New York: Routledge).

Forrester, James William. 1984. Gentle Murder, or the Adverbial Samaritan, *Journal of Philosophy*, **81**, 193–97.

Graham, Peter A. 2011. "Ought" and Ability, *Philosophical Review*, **120**, 337–82.

Graham, Peter A. 2016. A Degrees-of-Moral-Wrongness Argument That the Numbers Do Count, *Pea Soup*, <http://peasoup.us/2016/04/a-degrees-of-moral-wrongness-argument-that-the-numbers-do-count/> [accessed 6 April 2019].

Howard-Snyder, Frances. 2006. "Cannot" Implies "Not Ought," *Philosophical Studies*, **130**, 233–46.

Hurka, Tom. 2019. More Seriously Wrong, More Importantly Right, *Journal of the American Philosophical Association*, **5**, 41–58.

Kamm, F. M. 2007. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (New York: Oxford University Press).

Lazar, Seth. 2018. In Dubious Battle: Uncertainty and the Ethics of Killing, *Philosophical Studies*, **175**, 859–83.

Lazar, Seth & Lee-Stronach, Chad. 2019. Axiological Absolutism and Risk, *Noûs*, **53**, 97–113.

Lee-Stronach, Chad. 2018. Moral Priorities Under Risk, *Canadian Journal of Philosophy*, **48**, 793–811.

Lewis, David. 1974. Semantic Analyses for Dyadic Deontic Logic, in *Logical Theory and Semantic Analysis*, ed. by Sören Stenlund, Ann-Mari Henschen-Dahlquist, Lars Lindahl, and others (Dordrecht: Reidel, 1974), 1–14.

Louise, Jennie. 2004. Relativity of Value and the Consequentialist Umbrella, *The Philosophical Quarterly*, **54**, 518–36.

Oddie, Graham & Milne, Peter. 1991. Act and Value: Expectation and the Representability of Moral Theories, *Theoria*, **57**, 42–76.

Peterson, Martin. 2010. A Royal Road to Consequentialism? *Ethical Theory and Moral Practice*, **13**, 153–69.

Pettigrew, Richard. 2016. *Accuracy and the Laws of Credence* (Oxford: Oxford University Press).

Portmore, Douglas W. 2007. Consequentializing Moral Theories, *Pacific Philosophical Quarterly*, **88**, 39–73.

Portmore, Douglas W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press).

Potochnik, Angela. 2017. *Idealization and the Aims of Science* (Chicago: University of Chicago Press).

Prakken, Henry & Sergot, Marek. 1997. Dyadic Deontic Logic and Contrary-to-Duty Obligations, in *Defeasible Deontic Logic*, ed. by Donald Nute (Dordrecht: Springer Science+Business Media), 223–62.

Rancourt, Benjamin T. 2017. Better Understanding Through Falsehood, *Pacific Philosophical Quarterly*, **98**, 382–405.

Ryan, Sharon. 2003. Doxastic Compatibilism and the Ethics of Belief, *Philosophical Studies*, **114**, 47–79.

Schroeder, S. Andrew. 2017. Consequentializing and Its Consequences, *Philosophical Studies*, **174**, 1475–97.

Sinnott-Armstrong, Walter. 1984. "Ought" Conversationally Implies "Can," *The Philosophical Review*, **93**, 249–61.

Stocker, Michael. 1971. "Ought" and "Can," *Australasian Journal of Philosophy*, **49**, 303–16.

Styron, William. 1979. *Sophie's Choice* (New York: Random House).

Suikkanen, Jussi. 2019. Consequentializing Moral Dilemmas, *Journal of Moral Philosophy* (2019), Online first: <https://doi.org/10.1163/17455243-20182787>.

Talbot, Brian. 2016. The Best Argument for "Ought Implies Can" Is a Better Argument Against "Ought Implies Can," *Ergo, an Open Access Journal of Philosophy*, **3**, 377–402.

Vallentyne, Peter. 1989. Two Types of Moral Dilemmas, *Erkenntnis*, **30**, 301–18.