METAPHILOSOPHY

METAPHILOSOPHY
Vol. 50, Nos. 1–2, January 2019
0026-1068

THE COST OF CONSEQUENTIALIZATION

HANNO SAUER

Abstract: Consequentializers suggest that for all non-consequentialist moral theories, one can come up with a consequentialist counterpart that generates exactly the same deontic output as the original theory. Thus, all moral theories can be "consequentialized." This paper argues that this procedure, though technically feasible, deprives consequentialism of its potential for normative justification. By allowing purported counterexamples to any given consequentialist moral theory to be accommodated within that theory's account of value, consequentializers achieve a hollow victory. The resulting deontically equivalent consequentialist counterpart that results from absorbing originally non-consequentialist moral intuitions can now no longer explain, in a theoretically illuminating way, why certain actions are wrong and others right. The paper explains why traditional consequentialist theories did not embrace the procedure, and sketches how consequentialism can consequentialize without incurring the same cost.

Keywords: consequentialization, consequentialism, moral theory, theory of value.

Introduction

Normative ethicists are now all under the "consequentialist umbrella" (Louise 2004), some say. Some would rather get wet, and refuse to include consequentialists under it (Schroeder 2006; Sachs 2010). And some who are already under the umbrella do not want others to join them (Pettit 1997; Portmore 2007). In this short paper, I choose a different strategy and argue that the umbrella does not exist at all.

I take as my starting point Campbell Brown's recent paper (2011) on the limits of consequentialization. If, as I argue below, the consequentialization procedure as described by Brown and others comes with a significant cost, then his and other consequentializers' accounts are not really about the limits of consequentializing (or lack thereof) but simply about the limits of consequentialism *tout court*.

Here is how the argument goes. My paper has four sections: in the first section, I briefly and generally describe the consequentialization procedure. In the second section, I go into (some of) the details of Brown's formal model and show to what extent Brown's (mis)conception of the structure of consequentialism is reflected by this model. In the third

section, I explain what theoretical costs the incompleteness of his model comes with. Finally, in the fourth section, I show how to consequentialize properly—and conclude that this type of consequentialization is different from what the consequentialization procedure is usually taken to be.

To say that the consequentialist umbrella, at least as it is typically understood, does not exist does not, of course, entail that *consequentialism* as such does not exist or that it is incoherent. The consequentialist umbrella is not meant for consequentialism itself. Therefore, its non-existence does not mean consequentialism vanishes with it. Rather, the umbrella is for those moral theories that were previously thought to be paradigmatically non-consequentialist. By offering the umbrella to them, the consequentialist suggests that most, if not all, putatively non-consequentialist theories can be consequentialized. This is an offer that, I shall argue, has to be refused.

The Procedure

Many have attempted to inflict *the* fatal blow on consequentialism. Usually, refutations of consequentialism go like this: start with the theory itself (*C*); devise a counterexample (*E*) to *C* by showing that *C* entails that *E* is morally permissible; argue that this action clearly is not morally permissible (*E*). Conclude that *C* is false.

Simple enough. But wait: consequentializers hold that, as long as certain basic constraints are met, all such counterexamples can be accommodated within a consequentialist framework, because consequentialism as such can entail many things, depending on what theory of the good it is combined with. In order to consequentialize the theory that is not vulnerable to the same counterexample, and thereby render consequentialism itself immune to it, one simply has to come up with a theory of the good that explicitly accounts for the value the counterexample invokes as incompatible with *C*: "The main strategy for 'consequentializing' any given moral theory is simple. We merely take the features of an action that the theory considers to be relevant, and build them into the consequences" (Dreier 1993, 23). How are we supposed to go about doing this? Many non-consequentialists, for example, are bothered by the fact that some versions of consequentialism entail that breaking a promise can be permissible. This is what the consequentializer recommends: "[I]f a theory says that promises are not to be broken, then we restate this requirement: that a promise has been broken is a bad consequence" (Dreier 1993, 23). Or suppose the counterexample you find most compelling is that some versions of consequentialism would allow that some people be killed and their organs be harvested and distributed to a large number of terminally ill patients. Now the consequentializer will say that this is not a counterexample either, because nothing prevents the consequentialist from including the violations of people's rights among her bad consequences.

Here is a more thorough description of the strategy: "Take whatever considerations that the nonconsequentialist theory holds to be relevant to determining the deontic status of an action and insist that those considerations are relevant to determining the proper ranking of outcomes. In this way, the consequentialist can produce an ordering of outcomes that when combined with her criterion of rightness yields the same set of deontic verdicts that the nonconsequentialist theory yields—that is, for any deontic predicate ('permissible', 'impermissible', 'obligatory', 'supererogatory', etc.), the resulting consequentialist counterpart theory and the original nonconsequentialist theory will be in perfect agreement as to the set of actions that are in the extension of that predicate" (Portmore 2007, 39). Brown refers to the set of actions that, according to a moral theory *T*, are in the extension of the respective deontic predicates as the *deontic output* (2011, 755) of *T*. *T* and *C* will often have a different deontic output at first; but after one combines *C* with a suitably improved theory of value *V*, *C* + *V* and *T* can become *deontically equivalent*; when this happens, *T* has been consequentialized.

A Formal Model

A moral code is a rightness function that selects, for a particular agent, all and only those actions from a set of possible worlds which are permissible. Brown's notation for this is $R_i(A)$ (2011, 758ff.). *A* is a set of possible worlds from which the agent *i* can choose. If an agent has exactly two options to choose from, only one of which is morally right, then the function representing the respective moral code says that $R_i(\{w_1, w_2\}) = \{w_1\}$.

Brown says, following Railton (1984, 148), that "one has adopted no morality in particular even in adopting consequentialism unless one says what the good is" (2011, 754). It is of course correct that the consequentialist principle to do what would make things go best does not amount to a particular "morality" unless paired with a theory of the good—otherwise, it delivers no "deontic output" whatsoever. (In the above example, $\{w_1\}$ is *R*'s deontic output.)

What selection is made from this set of possible worlds then determines whether the theory is consequentialist. For example, if the rightness function, whatever it is, allows the following: $R_{me}(\{w_m, w_y\}) = \{w_m\}$ & $R_{you}(\{w_m, w_y\}) = \{w_y\}$, then it cannot be consequentialist, because it violates the requirement of agent neutrality: it tells one agent (me) to do something other than it tells another agent (you), even though our options are the same.

The problem is that as it stands, Brown's formal model allows not only evaluative focal points (Kagan 2000)—for example, actions—but also theories of value, which are supposed to specify *why* some things have value (for example, because they maximize pleasure or satisfy people's informed preferences), to be built into the set of worlds out of which the candidate

rightness function can select some as being permissible to be brought about. An evaluative focal point—or, as I shall refer to it, the *evaluandum* of a theory—is a set of things from which a moral theory selects some as permissible to have or perform. Possible *evaluanda* can be actions, motives, intentions, traits, and so forth. A theory of value—or, as I shall refer to it, the *evaluans* of a theory—is a general theory that explains, in a non-trivial way, why a certain selection is made rather than another. Possible *evaluanses* can be *performing actions whose maxims cannot be universalized without contradiction is wrong*, *pleasure is good and pain bad*, *God's will ought to be obeyed*, and so forth.

The main point I wish to argue for is that by allowing the things that are supposed to be evaluated by a theory of value to become part and parcel of that theory, the particular consequentialist theory loses the power to *justify* why some things are evaluated the way they are. Utilitarianism, for instance, is supposed to offer a normatively convincing and explanatorily potent account of *why* poetry is better than playing Push-pin, namely, because there are qualitative differences between types of pleasure. Such a normatively convincing explanation is rendered superfluous by simply building into one's list of bad consequences that preferring Push-pin over poetry is bad.

The Costs of Consequentialization

In a nutshell, my argument is that the consequentializer ought not to include particular action types—such as *keeping promises* or *respecting individual's rights*—in her theory of the good, because this makes it impossible to justify why said action types are among the things that are morally permissible to do.

A moral theory is a general framework that specifies not only which things are right or wrong but also why they are right or wrong. For a particular consequentialist theory to deliver any deontic output *and* a justification for why it is this output rather than another, one needs three elements, not two. The consequentialist needs:

1. The consequentialist principle: an *evaluandum* is right iff it will make things go best.
2. A specification of the evaluative focal point of the consequentialist principle: a particular type of *evaluandum*, the thing that is to be evaluated by the consequentialist principle, such as actions, rules, or motives.
3. A theory of what makes the *evaluandum* good and, ultimately, better than other *evaluanda*; that is, an *evaluans*: for example, a particular consequentialist theory could say that what makes things go best (the *consequentialist principle*) is determined by the rules (the

evaluandum) whose general acceptance best satisfies people's preferences or maximizes pleasure (the *evaluans*).

The account of consequentialism that consequentializers apparently have in mind (and is reflected by Brown's formal model), it seems to me, remains too coarse-grained to explicitly distinguish between (2) and (3), and the resulting characterization of consequentialism is thus incomplete. The appearance that any moral theory can be consequentialized results from the misleading assumption that particular *evaluanda* can contribute to a theory of the good just as much as particular *evaluanses* can. So, for example, if your consequentialist theory recommends breaking promises or breaching people's rights (by cutting them up and harvesting their organs) because it would maximize pleasure, it appears that you can simply build a provision against promise breaking or rights breaching into the theory of good which, in conjunction with the consequentialist principle, now yields the results you desire for their potential to immunize your consequentialist theory against the proposed counterexample.

But this appearance is wrong-headed, because descriptions of act types such as *keeping promises* or *respecting rights* are not the kind of thing that can be built into the theory of the good. They are merely particular specimens of *evaluanda* from a possible evaluative focal point. Only properties such as *maximizes pleasures* and *satisfies people's preferences* can be among the plausible candidates for a theory of the good, because they offer a justification for *why* not killing other people, helping those in need, keeping promises, and respecting people's rights are morally obligatory at all. This is what the theory of the good is supposed to do: offer a justification for why the rightness function of the moral theory at issue is the way it is.

What would happen if this third element were missing? Suppose the structure of consequentialist theories would allow that certain *evaluanda* taken from your evaluative focal point could be built into the theory of good that purportedly renders the consequentialist principle a full-blown moral theory. What would happen is that now the consequentialist could have *literally nothing* to say about why one ought to keep promises, other than that they are valuable. She could offer no justification in terms of a moral theory why one ought to keep one's promises anymore. But that is exactly what consequentialism as a normative theory is supposed to be able to do.

It is worth asking why Bentham, Mill, or Sidgwick did not accommodate certain intuitions about what morality requires by building (in Bentham's case) *humane punishment is good* or (in Mill's case) *free speech is valuable* into their respective theories of the good. They had independent theories of the good (for example, a hedonic calculus and qualitative welfarism) that were supposed to be able to show *why* humane punishment

or free speech are valuable and, if necessary, to rule out some of our common-sense intuitions as illegitimate—which, as a side note, has become impossible if anything can, simply by *fiat*, be made part of one's theory of value. This is a large theoretical cost that flows directly from adopting the consequentialization strategy as described above.

Now suppose that you have a version of objective hedonistic act consequentialism in front of you. The proponent of this theory tells you that murdering other people is morally forbidden. You ask him why. He responds, by referring to his theory of the good (the *evaluans*), that murdering other people (the *evaluandum* selected from the evaluative focal point "actions") is bad because it violates the requirement that one ought to maximize pleasure. Similarly, the proponent of a rule-centred version of preference consequentialism could, when asked about why one ought to keep one's promises, respond that keeping one's promises would be in the set rules that, if universally followed, would best satisfy people's preferences. *This* is the justification for why one ought to keep one's promises. A complete formal model should therefore look something like this: $T_r \rightarrow [(R_{me}(\{w_m, w_y\}) = \{w_m\} \ \& \ R_{you}(\{w_m, w_y\}) = \{w_y\})]$. (In this example, T_r stands for the agent-relative normative theory that yields the given deontic output.) *Keeping one's promises* is the kind of thing that can be among the possible things from which the rightness function can select the deontic output that determines the function's normative content. *Maximizing pleasure*, on the other hand, is the kind of thing that belongs to the justificatory part on the far left: it specifies, in theoretically fertile terms, why this particular selection is made. It can never be among the evaluated elements of the choice situation, because to maximize pleasure is not a concrete action type but an *evaluative property of action types*. It would be absurd to tell someone who wants to know what to do that he should choose the action *maximize pleasure*. Rather, actual choice situations contain descriptions of specific action types, and that is what people want to be advised about. Should I keep this promise or break it? This is a meaningful question. And a meaningful response would be, for instance: you should keep/break it, because that would be recommended by the best set of rules/maximizing pleasure.

How to Consequentialize

What I have intended to show is that the general strategy of consequentialization, as described by Brown, Dreier, or Louise comes at the cost of a loss of justificatory power, because it misconstrues the general internal structure of consequentialist moral theories.

In arguing this way, I realize that I am putting myself in an awkward dialectical position; a position, indeed, that might be thought to unfairly accuse the consequentializer of failing to do something she simply did not

want to do. Consequentializers say that consequentialism is the theory that one ought to maximize the good and that what counts as good is a list that, if necessary, can be expanded upon. The consequentializer and I are in agreement about this. Our disagreement is about *how* the expansion of this list is supposed to happen: the consequentializer holds that with some exceptions—such as, for example, moral dilemmas—virtually anything can be included in her theory of the good. I, on the other hand, wish to suggest that consequentialism has more internal structure.

It is of course true that one can understand consequentialism any way one pleases. Therefore, no knockdown argument can be based on a suggestion to understand it differently, and I do not aim to present such an argument. My aim is to point out the cost that comes with a particular understanding of consequentialism, and this cost, I have argued, is the loss of consequentialism's justificatory power. To preserve this power, one must impose further structural constraints on what can be included in one's theory of the good, and distinguish it from the things whose moral value is assessable by this theory. The need for this internal structure has been implicitly acknowledged by traditional consequentialists, because they intended their moral theories to be able to justify the demands of morality, and not merely say what these demands are.

It should be emphasized that for those who insist on agreeing with the consequentializer's understanding of the purpose of consequentialism this complaint will be of no interest whatsoever. Maybe, then, the merit of my discussion—if it has any—is that it points out which of the theoretical ambitions consequentialists used to have are sacrificed if one adopts the consequentializer's understanding of consequentialism. And when I say that there is no such thing as consequentialization, this should be taken to mean that if, and only if, one shares traditional consequentialism's justificatory ambitions, the consequentialization strategy that has become so fashionable recently makes little sense.

Now it could be that some moral theories, for instance, Kantianism, can be consequentialized if it can be shown that there is a version of subjective non-hedonistic rule consequentialism that is, as far as its deontic output is concerned, extensionally equivalent to Kantian moral theory. But this type of consequentialization can only be achieved indirectly: it cannot be achieved simply by building the verdicts the non-consequentialist theory yields into one's theory of the good in the hope that the resulting consequentialist counterpart will then have become extensionally indistinguishable. It can only be achieved by offering a revised theory of the good that, when applied to the evaluative focal points in question, shows that the newly constructed consequentialist theory would recommend the same action. This is what happens, for instance, when a transition from act consequentialism to rule consequentialism is made. Kantians could say that act consequentialism is false because it says that sometimes one ought to break one's promises for the sake of the greater good. Intuitively, this

does not seem right. But, rather than directly counting promise breaking among an action's evaluatively bad consequences, consequentialists can make a transition from act consequentialism to rule consequentialism to accommodate this counterexample, and perhaps this revised theory will now yield the same moral judgment that the Kantian theory did. But notice that this transition has not been made simply by saying that keeping promises is now among the things that are intrinsically valuable, which would be a dodgy move indeed. It has been made by proposing a new and better theory of consequentialism, one that manages to deliver plausible moral verdicts as well as a workable justification for why those verdicts ought to be accepted.

A moral theory can be consequentialized by showing that there is a consequentialist theory for which the same actions fall under the same deontic predicates. But moral theories or, to be more precise, the deontic output of certain putatively non-consequentialist theories can only be consequentialized in a *normatively and theoretically useful* way by one of the following means. The consequentialist can:

1. change her theory of what has value (for instance, by moving from a hedonistic theory of value to a desire-satisfaction theory),
2. change the way her theory of value translates into a theory of what has overall value (for instance, by moving from a maximizing to a satisficing theory),
3. propose a different or more comprehensive list of evaluative focal points (for instance, by moving from act consequentialism to rule consequentialism, by including expected rather than actual consequences, and so on, in her list of *evaluanda*).

The whole point of consequentialization is of course to make precisely the move mentioned in (1). But I have argued that the particular way the consequentializer understands this move is *not* theoretically useful, given certain justificatory ambitions. What cannot—or, rather, *ought not to*—be done is to include what should be among the candidates for evaluative focal points in the theory of what has value. That promise keeping and the protection of rights are among the things that ought to be considered valuable is supposed to *follow* from the more basic elements of the theory specified above, not be factored into its premises. Making this move would eradicate all the genuinely normative, justificatory power consequentialism is supposed to have.

Finally, consider an analogy. Think about our reply to someone with Kantian proclivities who argues that all moral theories can be deontologized. We can, this Kantian says, incorporate any counterexample to her moral standard (that for some maxim or principle to be morally permissible, it must be possible to think/will it to be a universal law) by changing

our theory of value in a way that is analogous to the one consequentializers recommend. Consequentializers hold that we can consequentialize putative counterexamples by amending our theory of value. In the case of consequentialism, this means that we add an item (for example, promise breaking) to our list of things that constitute a *bad consequence*. Since Kantians famously ground what is good in what is right rather than the other way round, to deontologize something would thus mean to make an analogous move by simply adding something to our list of what makes something *wrong*.

Suppose Kantians argue that lying is never morally permissible. Now someone presents a possible counterexample: a group of homicidal thugs are knocking on your door, in pursuit of an innocent stranger whom you are hiding in your attic. If lying is never permissible, then one may not lie even to this group of would-be murderers—obviously, this is too much to bear, so (this version of) Kantianism must be false. Not so fast, replies the Kantian; I shall simply deontologize your counterexample by adding "lying to a group of murderous thugs in pursuit of an innocent individual" to my list of things that are *permissible*, just as consequentializers take themselves to be entitled to add things to their list of things that are *bad*.

In this deontological case, this move would clearly strike us as unacceptably ad hoc. Why is this? The suggestions developed above are that we would want a deontological moral theory to show *why* certain acts are wrong on the basis of an independently justified theory of rightness—in this case, the theory that what makes an action permissible is that its maxim can be universalized without contradiction in willing or conception. This explanatory and justificatory step is missing in the deontologization method just described. Yet the analogous strategy is precisely the one advocated by consequentializers.

Utrecht University
Department of Philosophy and Religious Studies
Janskerkhof 13
3512 BL Utrecht
The Netherlands
h.c.sauer@uu.nl

Acknowledgments

I would like to thank an anonymous *Metaphilosophy* reviewer for helpful feedback.

References

Brown, C. 2011. "Consequentialize This." *Ethics* 121, no. 4:749–71.
Dreier, J. 1993. "Structures of Normative Theories." *Monist* 76:22–40.

Kagan, S. 2000. "Evaluative Focal Points," In *Morality, Rules and Consequences: A Critical Reader*, edited by B. Hooker, E. Mason, and D. E. Miller, 134–55. Edinburgh: Edinburgh University Press.
Louise, J. 2004. "Relativity of Value and the Consequentialist Umbrella." *Philosophical Quarterly* 54, no. 217:518–36.
Pettit, P. 1997. "The Consequentialist Perspective." In *Three Methods of Ethics: A Debate*, edited by M. Baron, P. Pettit, and M. Slote, 92–174. Malden, Mass.: Blackwell.
Portmore, D. 2007. "Consequentializing Moral Theories." *Pacific Philosophical Quarterly* 88:39–73.
Railton, P. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy Public Affairs* 13, no. 2:134–71.
Sachs, B. 2010. "Consequentialism's Double-Edged Sword." *Utilitas* 22, no. 3:258–71.
Schroeder, M. 2006. "Not So Promising After All: Evaluator-Relative Teleology and Common-Sense Morality." *Pacific Philosophical Quarterly* 87:348–56.