Consequentialism's Double-Edged Sword

BENJAMIN SACHS

Program in Environmental Studies and Center for Bioethics, New York University

Recent work on consequentialism has revealed it to be more flexible than previously thought. Consequentialists have shown how their theory can accommodate certain features with which it has long been considered incompatible, such as agent-centered constraints. This flexibility is usually thought to work in consequentialism's favor. I want to cast doubt on this assumption. I begin by putting forward the strongest statement of consequentialism's flexibility: the claim that, *whatever* set of intuitions the best non-consequentialist theory accommodates, we can construct a consequentialist theory that can do the same while still retaining whatever is compelling about consequentialism. I argue that if this is true then most likely the non-consequentialist theory with which we started will turn out to have that same compelling feature. So while this extreme flexibility, if indeed consequentialism has it (a question I leave to the side), makes consequentialism more appealing, it makes non-consequentialism more appealing too.

I. INTRODUCTION

Consequentialism has long been subject to criticism for failing to do as good a job as non-consequentialism of accommodating widespread intuitions about cases. Recent moral philosophers such as Hare, Sen, Scheffler, Railton, Brink and Broome have aimed to show that consequentialism has greater flexibility, or ability to accommodate these intuitions, than it has been credited for. But if we accept these widespread intuitions, why not just adopt non-consequentialism? The thought must be that we give up something important if we abandon consequentialism in favor of non-consequentialism. What I aim to show here is that consequentialists have not yet shown us how it can be true *both* (1) that consequentialism can deliver the intuitively plausible case-specific verdicts that non-consequentialism can deliver and (2) that we give up something important when we move from consequentialism to non-consequentialism. In fact, the very premises likely to be used in arguments for the first claim can be used to cast doubt on the second.

Before explaining why this is the case, I first need to make claims (1) and (2) more precise. In claim (1), the term 'the intuitively plausible case-specific verdicts that non-consequentialism can deliver' should be understood as designating *all* such verdicts. Or, more precisely it should be understood as designating the most intuitively plausible consistent *set* of case-specific verdicts (since there might be contradictory verdicts each of which we find intuitively plausible) that non-consequentialism can deliver, where 'case-specific verdicts' are

© Cambridge University Press 2010
Utilitas Vol. 22, No. 3, September 2010
doi:10.1017/S095382081000018X

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

claims about the moral permissibility of specific actions. Thus, claim (1) is the claim that consequentialism can deliver the most intuitively plausible set of claims about the moral permissibility of specific actions that non-consequentialism can deliver. This claim, which I will label 'Consequentializability', has been argued for by several authors.[1]

Claim (2) holds that we give up something important when we move from consequentialism to non-consequentialism. We need to be specific, however, about which versions of consequentialism and non-consequentialism we are concerned about. It seems that in considering whether to abandon consequentialism in favor of non-consequentialism, consequentialists should care specifically about whether the best version of consequentialism is better than the best version of non-consequentialism. Therefore, claim (2) should be construed as asserting that we give up something important in moving from the best version of consequentialism to the best version of non-consequentialism.

The remaining ambiguity stems from the presence of the term 'give up something important' in claim (2). What is it about the best version of consequentialism that consequentialists would not want to give up, but might have to give up in the move to the best version of non-consequentialism? Broadly speaking, there are two possibilities. First, there might be some compelling idea with which the best version of consequentialism is compatible and the best version of non-consequentialism is incompatible. Call this 'something important' the Compelling Idea. Second, the best version of consequentialism might embody some desirable quality, such as simplicity, power or coherence, that the best version of non-consequentialism lacks. Call this 'something important' the Desirable Quality. It is evident that the consequentialists who hold fast to claims (1) and (2) believe that (2) is true on the first interpretation, where the move from the best version of consequentialism to the best version of non-consequentialism requires giving up some Compelling Idea. When consequentialists argue for claim (1), they make ideas the issue; they implicitly accept the notion that normative ethical theories should be judged according to what ideas they can accommodate, and they proceed to argue that consequentialism does a better job of accommodating a certain group

[1] James Dreier, 'Structures of Normative Theories', *Monist* 76 (1993), pp. 19–30; Jennie Louise, 'Relativity of Value and the Consequentialist Umbrella', *Philosophical Quarterly* 54 (2004), pp. 518–36; Michael Smith, 'Neutral and Relative Value after Moore', *Ethics* 113 (2003), pp. 576–98; Douglas Portmore, 'Consequentializing Moral Theories', *Pacific Philosophical Quarterly* 88 (2007), pp. 39–73; Graham Oddie and Peter Milne, 'Act and Value: Expectation and Representability of Moral Theories', *Theoria* 57.1–2 (1991), pp. 42–56.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

of ideas – intuitively plausible case-specific verdicts – than we might have thought.

Given these two clarifications, we should now interpret claim (2) as saying that the best version of consequentialism is, and the best version of non-consequentialism is not, compatible with the Compelling Idea. This claim, which I will label 'Superiority', is reasonably imputable to Sen, Scheffler, and others, and has actually been argued for by Douglas Portmore.[2]

II. WHAT DO WE GET OUT OF THE TRUTH OF CONSEQUENTIALIZABILITY?

My goal is to demonstrate that the arguments likely to be used in support of Consequentializability can be used against Superiority. In fact, on the assumption that Consequentializability is true, I will make a prima-facie case for the denial of Superiority, which I will call 'Parity':

Parity: If the best version of consequentialism is compatible with the Compelling Idea, then the best version of non-consequentialism is compatible with the Compelling Idea.[3]

My argument will be lengthy and detailed, but the basic idea behind it is quite simple: since consequentialism tells us that the moral permissibility of an action depends on the value of the resulting state of affairs, what case-specific verdicts a version of consequentialism can deliver depends on what value theory it includes. If Consequentializability is true, that means that the consequentialist can construct a value theory such that her version of consequentialism delivers the most plausible set of case-specific verdicts that can be delivered in a non-consequentialist framework. And if the resulting version of consequentialism is compatible with the Compelling Idea, then a fortiori the value theory is too. Now take the version of non-consequentialism that delivers that same set of case-specific verdicts and call it N. In order to make N compatible with the Compelling Idea, thereby demonstrating the truth of Parity, we simply need to incorporate that same value theory into N. In other words, once the consequentialist shows how we can have the Compelling Idea and the most plausible set of case-specific verdicts in a single package, there is no more reason to think that the best version of non-consequentialism – the

[2] Douglas Portmore, 'Consequentializing Moral Theories', and 'Consequentializing', *Philosophy Compass* 4 (2009), pp. 329–47.

[3] The truth-functional form of Superiority is P & ∼Q (the best version of consequentialism is compatible with the Compelling Idea and the best version of non-consequentialism is not). P & ∼Q is equivalent to ∼(P→Q). The denial of ∼(P→Q) is P→Q, which is the truth-functional form of Parity (if the best version of consequentialism is compatible with the Compelling Idea then the best version of non-consequentialism is too).

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

version that delivers that set – will be incompatible with the Compelling Idea. All the non-consequentialist has to do in order to make her theory compatible with the Compelling Idea is accept the fruits of the consequentialist's labor.

This section will be dedicated to laying out this argument in a more formal manner.

We begin, again, by assuming that Consequentializability is true. To establish the truth of Consequentializability, one must take the most plausible set of case-specific verdicts that non-consequentialism can deliver and put forward a value theory that shows that these verdicts can be delivered in a consequentialist framework. Since my opponent and I agree, at least for the sake of argument, that this can be done, it is not necessary to illustrate fully *how* it can be done.[4] Nevertheless, a bit of illustration might help capture the flavor of what is going on.

Suppose, for instance, that the most plausible set of case-specific verdicts includes agent-centered restrictions. Agent-centered restrictions are restrictions on an agent's performing acts of a certain type. On a standard understanding of what consequentialism is, it appears not to be able to accommodate agent-centered restrictions. Consequentialists believe that an action token is permissible for an agent, A, if and only if there is no other action available to A that would have a better outcome. Given this claim, no matter how bad a type of action's outcome is it will be permissible (obligatory, in fact) to perform a token of that action if doing so will prevent an action of the same type from being undertaken multiple times. After all, if one person's X-ing leads to a bad outcome, then five people's X-ing must lead to a *really* bad outcome. Therefore, there are no types of actions such that all tokens of that type are impermissible for some agent.

Douglas Portmore has identified a value theory that gets consequentialists out of this bind. On this value theory, the value of the state of affairs that would result from an agent's X-ing is always relative to that person. In other words, value is agent-relative. Because of this, it might be impermissible for an agent to perform an action with a bad outcome even if performing the action would prevent five other people from doing the same. For instance, suppose the agent faces the choice of whether to kill an innocent person in order to prevent five other people from each killing an innocent person. We might reasonably argue that relative to the agent, it is better that he not engage in killing. Engaging in killing would be inconsistent with the inviolability of the person who gets killed. Since inviolability, if it exists, is grounded in personhood, one person's violability entails everyone else's violability.

[4] For an argument to the effect that it cannot be done, or at least that it is not likely to be accomplished, see Mark Schroeder, 'Teleology, Agent-Relative Value, and "Good"', *Ethics* 117 (2007), pp. 265–95.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

Since the agent has a reason to prefer that he be inviolable (this reason may be bound up in considerations of his dignity), he has a reason to prefer that others be inviolable, and thus that he not engage in killing. If the argument goes through, then the agent is required to refrain from killing. In other words, the agent is under an agent-centered restriction to not kill.

Although ultimately I am going to argue that the truth of Consequentialializability does not give us a reason to think that we give up something important in the move from consequentialism to non-consequentialism, I do want to concede that the truth of Consequentialializability would mean that it is possible to construct more plausible versions of consequentialism than was previously thought possible. Demonstrating how consequentialism can accommodate, for instance, agent-centered restrictions would indeed make consequentialism more appealing. I merely want to show that the truth of Consequentialializability would also, surprisingly, yield the result that we can construct more plausible versions of *non*-consequentialism than was previously thought possible.

The way I aim to explain this, again, is by taking Consequentialializability as given and arguing from it to a further thesis, Parity. I will argue for Parity by way of conditional proof. I will assume its antecedent (the claim that the best version of consequentialism is compatible with the Compelling Idea) and then conduct a search for a possible world in which its consequent (the claim that the best version of non-consequentialism is compatible with the Compelling Idea) is false. In other words, I will search for a counterexample to Parity. My conclusion will be that most likely there isn't one.

To begin the search for a counterexample, we first need to get some idea of what the Compelling Idea is. As a first pass we can say that the Compelling Idea either is or is not a proposition that is part of consequentialist theories themselves. If it is not, then it is going to have to be a metaethical claim. This is because in order for there to be a counterexample to Parity the Compelling Idea must be a proposition with which some normative ethical theories (including the best version of consequentialism) are compatible and others (i.e., the best version of non-consequentialism) are not. If the Compelling Idea is not itself a proposition about right and wrong, good and bad – the sort of proposition out of which normative ethical theories such as consequentialism are built – then it is going to have to be a proposition about metaphysics, epistemology or normativity that has some bearing on which propositions about right and wrong and good and bad are true. Otherwise, it will be irrelevant to the choice between consequentialism and non-consequentialism. Any metaphysical, epistemological or normative claim that is relevant to the identification of the correct normative ethical theory is, I shall stipulate, a metaethical claim.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

I am going to set aside the possibility that the Compelling Idea might be a metaethical claim, for the simple reason that the class of metaethical claims is just too large to be surveyed in the way we need. Given the great variety of metaethical claims that are out there, I cannot imagine what general considerations one might adduce to show that the Compelling Idea simply cannot be a metaethical claim. It bears mentioning, however, that it seems unlikely that the Compelling Idea will be a metaethical claim.

For one thing, some metaethical claims have no bearing on what normative ethical theories are true. Non-reductive realism is a good example. This is the view that moral properties are real features of the world that do not reduce to non-moral features of the world. We could accept or reject this as we wish and remain entirely neutral on substantive ethical matters. For another, the metaethical claims that do have substantive implications usually do not make the cut between acceptable and unacceptable at the line between consequentialism and non-consequentialism or at the line between some version(s) of consequentialism and everything else. For instance, in his latest defense of Consequentializability, Portmore identifies the following as the Compelling Idea:

> Teleological Conception of Reasons: The reasons there are for and against performing a given act are wholly determined by the reasons there are for and against preferring its outcome to those of its available alternatives, such that, if S has most reason to perform a*i*, then, of all the outcomes that S could bring about, S has most reason to desire that o*i* (i.e., a*i*'s outcome) obtains.[5]

Yet in order to connect the teleological conception of reasons to consequentialism, Portmore has to redefine consequentialism. According to him consequentialism is the view that

> an act's deontic status is wholly determined by the reasons there are for and against preferring its outcome to those of its available alternatives, such that, if S is morally required to perform a*i*, then, of all the outcomes that S could bring about, S has most reason to desire that o*i* (i.e., a*i*'s outcome) obtains.[6]

The other possibility, and hence the one more worthy of our attention for now, is that the Compelling Idea is a claim found in consequentialist theories. My goal is to search for a counterexample to Parity on the assumption that this is the case.

First, however, we need to identify the elements of consequentialist theories. I will be generous here, allowing in some kinds of claim that one might doubt are really parts of consequentialist theories themselves. (Given my argumentative strategy, this generosity can

[5] Portmore, 'Consequentializing', p. 333.
[6] Portmore, 'Consequentializing', p. 333.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

only help my opponent.) I am going to allow that consequentialist theories have four elements: First, one or more moral principles. Second, evaluative claims – claims about what is good. Moral principles, combined with non-moral facts and evaluative claims, yield the third part of a normative ethical theory: a set of case-specific verdicts. The fourth part is one or more determination claims – claims about what kinds of fact determine the truth of the case-specific verdicts. Determination claims are to be distinguished from moral principles in the following way: whereas moral principles state necessary and/or sufficient conditions for an action's rightness or wrongness, determination claims identify the kinds of fact that *ground* an action's rightness or wrongness.[7]

Having identified the elements of consequentialist theories, we can begin our search for a counterexample to Parity. The following notation will be used: The best non-consequentialist theory will be labeled 'N', the best consequentialist theory will be labeled 'C', and the most plausible set of case-specific verdicts that non-consequentialism can deliver will be labeled 'Δ'.

1. Moral principles

Suppose C's Compelling Idea is one of its moral principles. (This was Portmore's claim in his original defense of Superiority – he believed that the Compelling Idea was the proposition that it is always permissible to maximize the good.)[8] In virtue of what might N be incompatible with the Compelling Idea? Not in virtue of N's case-specific verdicts. By definition, Δ is N's set of case-specific verdicts. (Since N is defined as the best non-consequentialist theory and the best non-consequentialist theory would yield the most plausible set of deontic verdicts that a non-consequentialist theory can yield.) If Consequentializability holds, then Δ is also C's set of case-specific verdicts, and therefore Δ is compatible with the Compelling Idea.

The only other possibility is that N might be incompatible with the Compelling Idea in virtue of its own moral principles. But this turns out not to be true either. Consequentialist moral principles come in varied forms, but one generalization we can make about them is that they all assert some connection between evaluative facts and moral facts. Therefore, constructing N so as to make it compatible with a consequentialist moral principle is as simple as ensuring that N contains no principles asserting some connection between evaluative and moral facts.

[7] Shelly Kagan nicely elucidates the importance of moral explanation in *The Limits of Morality* (Oxford, 1989), pp. 13–15.

[8] 'Consequentializing Moral Theories'.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

No matter what appears in Δ, this will always be possible. There will usually be a number of ways to do this, but at least one method will always be available: the construction of a ten commandments-style list of DOs and DON'Ts, one for every member of Δ. For instance, if 'it is impermissible to kill Jones' is a member of Δ, just put 'DON'T kill Jones' on the list of principles in N. Since commandment-style principles make no claims about the connection between evaluative facts and moral facts, they are consistent with all consequentialist moral principles, including whichever one the Compelling Idea is.

One might object that N cannot contain a set of ten commandments-style principles since N, by definition, is the best version of non-consequentialism and the best version of non-consequentialism would not contain a set of ten commandments-style principles. After all, ten commandments-style theories, as Shelly Kagan has pointed out, lack simplicity, power and coherence.[9] Even if this is true, however, it does not help us find a counterexample to Parity. If N must be simple, powerful and coherent, then C must as well. However, I am not aware of any argument to the effect that there is a version of consequentialism that embodies these virtues and yields Δ. If there is no such version of consequentialism, then the antecedent of Parity is false, and thus Parity is trivially true.

Indeed, there is a strong reason to think that C does *not* possess these virtues: value theories that are simple, powerful and coherent do not, when combined with consequentialist moral principles, yield the kind of case-specific verdicts likely to appear in Δ, and any normative ethical theory that takes on a complex, weak and non-coherent value theory in order to accommodate these verdicts is ipso facto complex, weak and non-coherent. Portmore, for instance, in constructing a version of consequentialism that accommodates Δ, puts forward a value theory with a host of odd features, including agent-relative value, dual ranking (ranking of outcomes on two independent criteria) and arbitrary multiplicative weights ('Consequentializing Moral Theories', Section IV). That Portmore's version of C fails to embody the three virtues is, of course, insufficient to establish that C *must* fail on this count, but then neither have we been shown that N must have this defect.

2. *Evaluative claims*

Suppose C's Compelling Idea is one of its evaluative claims. As I just argued, on the assumption that Consequentializability is true we can construct N such that it delivers Δ and contains no moral principle asserting a connection between evaluative facts and moral facts. This

[9] Kagan, *The Limits of Morality*, p. 11.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

being the case, we can append any value theory we wish to N. So no matter which evaluative claim the Compelling Idea is, we can make N consistent with it. Hence there is no counterexample to Parity to be found here.

3. *Case-specific verdicts*

Suppose that C's Compelling Idea is one of the case-specific verdicts that it yields. Again, by definition, Δ is N's set of case-specific verdicts. And if Consequentializability holds, then Δ is also C's set of case-specific verdicts. So if the Compelling Idea is a case-specific verdict then it will be found in Δ, and hence N will yield it. Hence there is no counterexample to Parity to be found here either.

4. *Determination claims*

I argued above that if the Compelling Idea is a moral principle, then, assuming Consequentializability holds, we will be able to construct N such that the Compelling Idea is compatible with it. Thus I am relying heavily on a heretofore implicit claim that non-consequentialist theories can contain consequentialist moral principles. My view, which I take to be intuitively plausible enough not to require defense, is that a normative ethical theory that accepts a mixture of consequentialist and non-consequentialist moral principles is non-consequentialist if its determination claims are non-consequentialist. A determination claim is non-consequentialist just in case it holds that the kinds of fact that determine the permissibility of an action are not exclusively facts about the goodness of the state of affairs that the action would bring about (or, for rule-consequentialism, facts about whether the action is prohibited by the set of rules the general adherence to which would be most conducive to the common good).[10]

Imagine, for instance, a theory that includes the following two moral principles:

(a) An act is wrong if its performance under the circumstances would be disallowed by any system of rules for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement.
(b) An act is wrong if there was an alternative to it available to the agent that would have brought about a better state of affairs.

Such a theory is non-consequentialist if it also contains the following determination claim:

[10] We should of course construe 'state of affairs' broadly enough to encompass facts about the act itself. That way a determination claim remains consequentialist even if it takes into account, for instance, the badness of a right being violated.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

(c) Every wrong action is wrong because its performance under the circumstances would be disallowed by any system of rules for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement.

Being lenient about moral principles requires being strict about determination claims. To defend Parity while insisting that non-consequentialist theories can accept consequentialist moral principles, we are going to have to say that for a theory to be non-consequentialist it must contain a non-consequentialist determination claim. And this raises the prospect that if C's Compelling Idea is a determination claim then we will not be able to construct N such that it is compatible with that claim. This prospect is made all the more likely by the fact that determination claims are by their very nature exclusive. To illustrate the point, consider the following determination claim:

(d) Every wrong action is wrong because there was an alternative to it available to the agent that would have brought about a better state of affairs.

We have to say that (c) and (d) are incompatible unless we want to admit the possibility of widespread moral overdetermination. By contrast, there is nothing troublesome about saying that (a) and (b), the two moral principles, are compatible.

So if C's Compelling Idea is a determination claim, then we have our counterexample to Parity. But I doubt that C's Compelling Idea is a determination claim. On the contrary, in fact, consequentialism's determination claims are usually seen as one its main drawbacks. If Jones stalks and kills an innocent person in cold blood, any consequentialist theory will tell us something like: it's the resulting state of affairs that grounds the rightness or wrongness of Jones's action. Thus, even if consequentialism gives the right answer in this case – that Jones acted wrongly – it will give it for what is intuitively the wrong reason. This is a well-worn complaint, of course,[11] but strikingly few theorists have made an effort to respond to it[12] or to give an argument for a consequentialist determination claim.[13]

[11] See, for instance: John Rawls, *A Theory of Justice* (Cambridge, Mass., 1971), pp. 26, 28; Bernard Williams, 'A Critique of Utilitarianism', in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge, 1973), p. 99; Tim Mulgan, 'Rule Consequentialism and Famine', *Analysis* 54 (1994), pp. 187–92.

[12] An exception is William Shaw, 'The Consequentialist Perspective', *Contemporary Debates in Moral Theory*, ed. J. Dreier (Malden, Mass., 2006).

[13] Part of the problem is that the distinction between consequentialism's moral principles claims and its determination claims often goes unnoticed, and therefore some theorists think they are defending the latter when in fact they are defending the former. See, for instance, R. Euguene Bales, 'Act-Utilitarianism: Account of Right-Making

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

Now one might object that I have unfairly stacked the deck against consequentialist determination claims by presenting an insufficiently general version of the one we have been discussing. I said just now that it seems that consequentialism will give the right answer for the wrong reason if it says that Jones's cold-blooded killing is wrong on the grounds that Jones could have brought about a better state of affairs by doing something else. While such a determination claim looks less than compelling, we might think that the general claim of which it is an instance is unimpeachable. That claim would be something along the lines of 'morality is about doing what's best'. One might even suppose that what accounts for the literature's dearth of defenses of consequentialist determination claims is that such claims, when abstracted to a sufficient level of generality, are just obviously true.

This appearance, I submit, is deceptive. Notice, first, that 'morality is about doing what's best' is not a determination claim. A determination claim identifies some facts as determining the truth of some proposition(s). 'Morality is about doing what's best' doesn't do this. The determination claim that best captures the spirit of this proclamation is, 'The fact whether X-ing will bring about the best available state of affairs determines whether X-ing is permissible.' Already this begins to look more debatable.

And things get worse for the consequentialist, for her determination claim, as we are now interpreting it, is ambiguous in a troubling way. There is an ambiguity in the term 'best' – the question being whether we are talking about moral or non-moral bestness. Suppose it's the former. In this case, the consequentialist determination claim would read: the fact whether X-ing will bring about the morally best state of affairs determines whether X-ing is permissible. But a consequentialist cannot say this, for the following reason. If she believes in moral bestness, then she believes in moral goodness. And if anything is morally good, surely a permissible act is, qua permissible act. Or, at the very least, permissible acts are morally better than impermissible ones. Therefore, whether X-ing would bring about the morally best possible state of affairs depends on whether X-ing itself is morally good. But if the determination claim we are considering were true, then it would also be the case that whether X-ing itself is morally good depends on whether X-ing would bring about the morally best possible state of affairs. Hence, this version of consequentialism entails circular explanation, in which A explains B and B explains A.[14]

Characteristics or Decision-Making Procedure?', *American Philosophical Quarterly* 8 (1971), pp. 257–65.

[14] William K. Frankena makes roughly the same point in *Ethics*, 2nd edn. (Englewood Cliffs, NJ, 1973), p. 14.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

Therefore the consequentialist must instruct us to interpret 'best' in her determination claim as signifying 'non-morally best'. So her determination claim is now, 'The fact whether X-ing will bring about the non-morally best state of affairs available to the agent determines whether X-ing is permissible.' This is not very compelling.

Thus, whether we look at consequentialist determination claims from afar or in an applied context, they don't look like the sort of thing that deserves the 'Compelling Idea' label. Therefore, we can admit that there would be a counterexample to Parity if the Compelling Idea were a determination claim, and yet be reasonably confident that there is no counterexample to Parity.

Taking the truth of Consequentializability as given, I have argued in this section in support of the claim, which I labeled Parity, that if the best version of consequentialism is compatible with the Compelling Idea, then the best version of non-consequentialism is as well. I have focused my efforts on determining just what shape the Compelling Idea might take such that the best non-consequentialist theory, N, could not accommodate it. I first dismissed the possibility that the Compelling Idea might be a metaethical claim. I then identified dispositive reasons to believe that N will not be incompatible with the Compelling Idea if the latter is a moral principle, evaluative claim or case-specific verdict. And we found a strong reason to think that the Compelling Idea is not a determination claim, meaning that we needn't worry about which determination claims N is or is not compatible with. Thus, we have a prima-facie case for the truth of Parity, and hence the falsity of Superiority. The burden of proof is now on the advocate of Superiority to undermine this case by showing one of the following two things: (1) the Compelling Idea is a consequentialist determination claim, or (2) the Compelling Idea is a metaethical claim with which the best version of consequentialism is compatible and the best version of non-consequentialism is incompatible.

III. CONCLUSION

Various theorists have argued for Consequentializability – the claim that consequentialism is flexible enough to deliver all the case-specific verdicts that non-consequentialism can deliver. I have argued that this extreme flexibility would turn out to cut both ways for consequentialism. I did this by assuming that consequentialism is indeed that flexible and then making a prima-facie case for believing that if the best version of consequentialism delivers these case-specific verdicts while remaining compatible with consequentialism's Compelling Idea, then the best version of non-consequentialism does the same. My case was based on pointing out that if Consequentializability is true then

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

non-consequentialists needn't deny any consequentialist claims except their determination claim, which is probably not consequentialism's Compelling Idea. On reflection, this result is not surprising. One cannot bring consequentialism closer to non-consequentialism without bringing non-consequentialism closer to consequentialism.

And what of those theorists who don't go so far as to defend Consequentializability – the theorists who argue that consequentialism is more flexible than it has been credited for, but do not depict consequentialism as flexible enough to accommodate all the case-specific intuitions that non-consequentialism can accommodate? This group includes many of the most prominent consequentialists, several of whom I mentioned at the outset: Hare, Scheffler, Sen, Railton, Brink and Broome. Since they do not argue for anything so extreme as Consequentializability, their arguments do not show that the most plausible version of non-consequentialism is compatible with consequentialism's Compelling Idea. And so their versions of consequentialism retain an advantage over non-consequentialism. But this advantage comes with a disadvantage, which is of course that their versions of consequentialism, by hypothesis, do not accommodate all the case-specific intuitions that non-consequentialism can. Such consequentialists are free to argue that the intuitions that they cannot accommodate are ones that (for independent reasons) we ought to reject. But if they are right that we ought to reject them, then the best non-consequentialist theory will reject them too. This version of non-consequentialism will issue the same set of case-specific verdicts as the best version of consequentialism, and so that version of non-consequentialism, for reasons explained in Section II, will most likely be compatible with consequentialism's Compelling Idea. Therefore, these consequentialists are in trouble either way.

What strategy, then, should consequentialists adopt in their effort to demonstrate the superiority of their theory? The lesson to be learned from the failure of the truth of Consequentializability to advance the cause of consequentialism is that consequentialists should not try to meet non-consequentialism on its own terms. They should in fact do the exact opposite. A completely unapologetic version of consequentialism, a version of consequentialism that makes no effort to accommodate widespread case-specific intuitions, is the consequentialist's best hope. Consequentialists should embrace unapologetic consequentialism – taking theorists such as Peter Singer, Shelley Kagan and Peter Unger as their model – and then either argue us out of our intuitions or convince us that our intuitions have no evidential force. The reason they should take this extreme step is that it is the only way to force non-consequentialists into an uncomfortable position. When consequentialists attempt to tweak consequentialism so that it

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press

can accommodate the case-specific verdicts that non-consequentialism has always been able to deliver, they have to insert all manner of strange epicycles, such as agent-relative value, into their value theories. But if instead consequentialists convince us that we should accept the case-specific verdicts that appear to follow from the classic versions of consequentialism, such as Bentham's or Sidgwick's, then it will be the *non*-consequentialists who have some explaining do to. Consequentialism's determination claim would appear to be a more simple, powerful and elegant explanation of the truth of those verdicts than any possible non-consequentialist determination claim. Consequentialists would then be well on their way to demonstrating that the best version of consequentialism embodies a desirable quality that the best version of non-consequentialism lacks.

This brings us back to the first fork in the road for the consequentialist who seeks to show that the best version of consequentialism is superior to the best version of non-consequentialism. That fork was the decision whether to insist that the best version of non-consequentialism fails to accommodate some Compelling Idea or that the best version of non-consequentialism fails to embody some Desirable Quality. I have just now suggested that the latter path is open to consequentialists. The former path, however, was always a dead end. The consequentialist gains little by showing that non-consequentialism fails to accommodate some Compelling Idea unless she can demonstrate that the various compelling ideas that non-consequentialism can accommodate – i.e., our case-specific intuitions – can be accommodated by consequentialism. But it turns out that the weapon consequentialists can wield in order to prove this latter claim undermines the arguments they might make for the former claim. The reason for this should be clear by now. Their weapon – consequentialism's flexibility – is a double-edged sword.[15]

ssachsben@aol.com

[15] This article was presented as a paper before audiences at the Department of Bioethics in the Warren G. Magnuson Clinical Center of the National Institutes of Health and at the 2008 annual meetings of the Wisconsin Philosophical Association and the International Society for Utilitarian Studies. Mark Schroeder, Jenny Louise, James Dreier, Joe Millum and Carla Saenz each read a previous draft of this article and aided me enormously with their comments. Finally, I owe a tremendous debt to Douglas Portmore for graciously helping me to straighten out my thoughts on this issue, correcting numerous mistakes, and offering encouragement on the worthiness of the project.

https://doi.org/10.1017/S095382081000018X Published online by Cambridge University Press