

Consequentialise This

Campbell Brown
University of Edinburgh

Draft revised 8th May 2009
Not for citation

1. Introduction

Arguing with a consequentialist can be frustrating. Witness a typical sort of exchange:

You—a non-consequentialist, let's assume—begin with your favourite counterexample. You describe some action—say, a judge's convicting an innocent man to avert a riot, or a doctor's murdering a healthy patient for her organs—which, you say, would clearly have the best consequences, yet equally clearly would be greatly immoral. So consequentialism is false, you conclude; sometimes a person ought not to do what would have best consequences.

'Not so fast,' comes the consequentialist's reply. 'Your story presupposes a certain account of what makes consequences better or worse, a certain theory of the good, as we consequentialists like to say. Consequentialism, however, is not wedded to any such theory of the good. We already knew that combining consequentialism with some theories of the good would have implausible results; that's what utilitarianism has taught us. In order to reconcile consequentialism with the view that this action you've described is wrong, we need only to find an appropriate theory of the good, one according to which the consequences of this action would not be best. You say you're concerned about the guy's rights?

No worries; we'll just build that into your theory of the good. Then you can be a consequentialist too.'

And just like that, you've been assimilated.¹ It's a cunning ploy. Instead of showing that your non-consequentialism is mistaken, the consequentialist shows that it's not really non-consequentialism; instead of refuting your view, he 'consequentialises' it. If you can't beat 'em, join 'em. Better still, make 'em join you.

One may suspect, however, that this consequentialising strategy, if used in excess, will ultimately undermine consequentialism. It might succeed in immunising consequentialism against counterexamples only at the cost of severely weakening it, perhaps to the point of utter triviality. So effortlessly is the strategy deployed that some are led to speculate that it is without theoretical limits: *every* moral view may be dressed up in consequentialist clothing.² Take any theory traditionally thought of as opposed to consequentialism, and it will be possible, they predict, to devise a consequentialist 'counterpart theory' which mimics perfectly the verdicts of the original regarding the moral status of actions, whether they are right or wrong, permitted or forbidden.³ If so, then consequentialism has something to offer everyone. No matter your ethical persuasion, there's a consequentialist theory just right for you. But then, it seems, consequentialism would be empty—trivial, vacuous, without substantive content, a mere tautology. The statement that an action is right iff it maximises the good would entail nothing more substantive than the statement that an action is right iff it is right; true perhaps, but not of much use.

As I aim to show here, however, this speculation is overblown. There

1. The Star Trek analogy is apt. Consequentialists are generally very cold and calculating, much like the Borg, and they're wont to ignore the separateness of persons.

2. Among the 'consequentialisers' who make this claim, or something close to it, we may include James Dreier, 'Structures of Normative Theories', *The Monist* 76 (1993): 22–40; Jennie Louise, 'Relativity of Value and the Consequentialist Umbrella', *The Philosophical Quarterly* 54 (2004): 518–536; Douglas W. Portmore, 'Consequentializing Moral Theories', *Pacific Philosophical Quarterly*, no. 88 (2007): 39–73.

3. This notion of a 'counterpart theory' is taken from Louise, 'Relativity of Value and the Consequentialist Umbrella'.

are in fact limits to consequentialisation. Some views of right and wrong are incompatible with consequentialism, no matter what theory of the good one adopts. These views cannot be consequentialised.

I'll proceed as follows. In Section 2, I'll say more precisely what consequentialisation entails. This of course will involve saying more precisely what consequentialism is. My intention will not be to give the one true, uniquely correct definition of consequentialism. Indeed I doubt there is any such thing; 'consequentialism' is a term of art used by philosophers to mean different things on different occasions, none of which is obviously most deserving of the name. Rather, my approach will be to adopt the 'best case scenario' for the consequentialiser. I'll try to define consequentialism in a way that makes consequentialisation as easy as possible. Then, in Section 3, I'll show that, even given this favourable definition of consequentialism, the project of consequentialising all theories cannot succeed. In particular, I'll distinguish three familiar sorts of moral theories which resist consequentialisation.

2. How to Consequentialise

2.1. Maximising the good

I begin with the familiar characterisation of consequentialism as the view according to which an action is right, or permissible, iff it *maximises the good*. What exactly does this mean?

You might say: by 'the good' is meant the class of all things good, the extension of the predicate 'good'; so to maximise the good is to make as many good things as possible. But this is implausible. Suppose you can either do *x*, producing one good thing, or do *y*, producing two. Does it follow that you will maximise the good by doing *y*? Not necessarily; for it might be that the one thing is very good, yet the other two things only moderately good, so that overall the one is more good than the other two combined.

The problem with the first answer, you might say, is that it treats being good as an all-or-nothing affair, when in fact it is more a matter of degree.

Not all good things are equally good; some are better than others. This suggests another answer. To maximise the good is to make things as good as possible, the best they can be. But how are we to understand the extent to which things are good when there are more things than one and they're good to varying extents? Some procedure such as the following might seem to hold the answer. First partition the world into a set of mutually exclusive and jointly exhaustive things (to avoid double-counting). Then assign to each thing in the partition a number representing the degree to which it is good. Finally add up all the numbers. This sum will represent the extent to which things are good overall.

Perhaps this procedure can be made to work. But it rests on a controversial assumption, which the consequentialist need not accept. This assumption, sometimes called 'atomism', is that when a whole is composed of parts, the extent of goodness of the whole must equal the sum of the extents of goodness of the parts (at least when the parts are exclusive and exhaustive).⁴ This, of course, is just what G. E. Moore warned us not to assume, and for good reason.⁵ Recall the earlier example: doing *x* will produce one good thing; doing *y* will produce two. But suppose now that each of the latter two things is, considered by itself, better than the former. Does it follow that you will maximise the good by doing *y*? Again, not necessarily; for it might be that, although the two things are each better in isolation, when combined they are worse.

We should resist building atomism into consequentialism. Instead, I suggest, we should understand the extent to which things are good, more holistically, as the extent to which the world as a whole is good. As David Lewis puts it, '[t]he way things are, at its most inclusive, means the way this entire world is.'⁶ To maximise the good, then, is to make the world, the sum of all things, as good as it can be; it is to act so as to bring about the best possible world (of those which can be brought about).⁷

4. Cf. Campbell Brown, 'Two Kinds of Holism about Values', *The Philosophical Quarterly* (2007).

5. G. E. Moore, *Principia Ethica* (Cambridge University Press, 1903), 28.

6. David Lewis, *On the Plurality of Worlds* (Blackwell, 1986), 1.

7. Cf. Fred Feldman, *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*

More precisely, I'll assume the following general picture. When a person acts, she faces a range of alternatives, different actions she might perform. For each action, I assume, there's a unique possible world that would obtain if the action were performed.⁸ We can call this world the *outcome* of the action. In effect, the person faces a choice between possible worlds. Each available action is the bringing about of a possible world, its outcome the possible world thereby brought about. Consequentialism then says an action is right iff its outcome is best.

2.2. Some Clarifications

Before moving on, three points of clarification are in order. First, consequentialism is here to be understood as a 'criterion of rightness', as it's commonly put, rather than as a 'decision procedure'.⁹ That is, it need not recommend that agents aim directly at maximising the good when deciding what to do. Plausibly, some outcomes, e.g. getting to sleep, are less well achieved by aiming directly at them, and maximising the good may be self-defeating in this way.¹⁰ In that case, consequentialism, as I'm thinking of it here, would recommend that agents decide what to do in some other way, e.g. by adopting 'rules of thumb'.¹¹

Second, consequentialism, as here understood, is a *purely extensional* thesis. It claims only that, among possible actions, certain properties, viz. those of being morally right and of maximising the good, are necessarily coinstantiated. It makes no claim of causation, determination, or explanation; it doesn't say that acts are right *because* their outcomes are best. Nor (if this amounts to a distinct claim) does it make a claim of priority;

(Cambridge University Press, 1997), 152.

8. This assumption is sometimes called 'counterfactual determinism'; see e.g. Krister Bykvist, 'Normative Supervenience and Consequentialism', *Utilitas* (2003): 30.

9. Cf. Philip Pettit and Geoffrey Brennan, 'Restrictive Consequentialism', *Australasian Journal of Philosophy* 64 (1986): 438–455; David Brink, 'Utilitarian Morality and the Personal Point of View', *The Journal of Philosophy* 83, no. 8 (1986): 417–438.

10. Cf. Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984), ch. 4.

11. Cf. J. J. C. Smart, 'Extreme and Restricted Utilitarianism', *Philosophical Quarterly* 6 (1956): 344–54.

it doesn't say that the right is *prior* to the good.¹² Sometimes it is objected that, while the extensional thesis may be all right, consequentialism puts the cart before the horse in one or other of these ways.¹³ Since, as I've said, I want to make things as easy as possible for the consequentialiser, I here understand consequentialism extensionally, so such objections may be set aside.

Finally, since the outcome of an action is here assumed to be an entire world, a person cannot in general know what the outcomes of her actions will be. Even assuming there is a fact of the matter which world would obtain should you act in a certain way, it's exceedingly unlikely that this fact would be accessible to you. Rather, in the normal case, you'll have various partial beliefs regarding which worlds might obtain conditional upon your action. In decision theory, such beliefs are represented by 'subjective probabilities' which figure in calculations of expected desirability. Thus, I'll here be considering only what is sometimes called *objective rightness*, which depends on what in fact the outcomes of actions are, and not *subjective rightness*, which depends on the variably probable outcomes of actions, as given by the agent's subjective probabilities. This I do solely in the interests of simplicity.

2.3. Theories of the Good

As Peter Railton explains, 'one has not adopted any particular morality in adopting consequentialism unless one says what the good is'.¹⁴ By itself, consequentialism is not a particular moral theory, but it becomes one, or a part of one, when combined with a theory of the good, a theory about which worlds are better or worse than which others. There are many theories of the good with which it might be combined; so there are many particular consequentialist theories. These are unified in holding

12. On the 'priority of the right' see John Rawls, *A Theory of Justice* (Oxford, 1971), 31.

13. See e.g. Elizabeth Anderson, 'Reasons, Attitudes, and Values: Replies to Sturgeon and Piper', *Ethics* 106 (1996): 538–554.

14. Peter Railton, 'Alienation, Consequentialism, and Morality', *Philosophy and Public Affairs* (1984): 148.

that we ought to maximise the good, but divided over what the good is. Consequentialism is the collection of all these particular theories taken together, the ‘genus’, we might say, of which they are ‘species’.

The most famous and oldest of these species is surely utilitarianism, which results from combining consequentialism with a hedonistic theory of the good. As contemporary consequentialists are quick to point out, however, consequentialism is not limited to utilitarianism. There is, they say, a tendency to conflate the two, which must be resisted, because it makes consequentialism appear in worse shape than it really is. Once it is divorced from the crude, naïve hedonistic theory of the good, many common objections evaporate. Railton, for example, gestures at a ‘pluralistic’ theory of the good in which ‘several goods are viewed as intrinsically non-morally valuable—such as happiness, knowledge, purposeful activity, autonomy, solidarity, respect, and beauty’.¹⁵ Such a theory, he argues, enables the consequentialist to escape the problems of ‘alienation’ that beset utilitarianism.

This flexibility of consequentialism as regards theories of the good opens the door to the consequentialisation. It raises the possibility that, by careful selection of an appropriate theory of the good, even the most adamant opponents of consequentialism may be brought under what Jenny Louise calls ‘the consequentialist umbrella’.¹⁶ James Dreier describes the strategy:

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. For example, if 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. Notice that the weighting is not yet specified. If the theory under consid-

eration includes an absolute side constraint against promise-breaking, then we have the consequentialist version give a lexically prior negative weight to promise-breaking.¹⁷

It will be helpful to describe the strategy more precisely. Let *M* be any moral theory which we aim to consequentialise. And let *C* be consequentialism. So a particular consequentialist theory is equivalent to some conjunction *C*&*G*, where *G* is a theory of the good. Say that the *deontic output* of a theory *P* is the set of all propositions about the rightness or wrongness of actions which are entailed by *P*; and say *P* and *Q* are deontically equivalent iff they have the same deontic output.¹⁸ Then to consequentialise *M*, we simply find some theory of the good *G* such that *M* is deontically equivalent to the particular consequentialist theory *C*&*G*. We might then call *C*&*G* a *consequentialist counterpart* of *M*.

But why, you might ask, should we think that this process of constructing a consequentialist counterpart for *M* amounts to the same thing as consequentialising it? Consider an analogy. Suppose I employ you to paint my red house blue, and you then proceed to build another house which is a perfect replica of mine except it’s painted blue rather than red. Then you haven’t done your job. The new house may be blue, but it’s not my house. You haven’t done anything to my house; it’s still red. Creating a blue counterpart of my house is not a way of making my house blue. Likewise, you might say, creating a consequentialist counterpart of *M* is not a way of making *M* a consequentialist theory. The counterpart theory, *C*&*G*, may be consequentialist, but that’s a different theory; *C*&*G* is not *M*. Of course the two theories do agree on all deontic questions, about which actions are right or wrong; by hypothesis, they’re deontically equivalent. But they needn’t agree on questions of the good, because *M* need not imply *G*, the theory of the good which is implied by *C*&*G*.

17. Dreier, ‘Structures of Normative Theories’, 23.

18. Cf. the ‘Deontic Equivalence Thesis’ in Portmore, ‘Consequentializing Moral Theories’.

15. Railton, ‘Alienation, Consequentialism, and Morality’, p. 149.

16. Louise, ‘Relativity of Value and the Consequentialist Umbrella’; Cf. the ‘consequentialist vacuum cleaner’ in David McNaughton and Piers Rawling, ‘Agent-Relativity and the Doing-Happening Distinction’, *Philosophical Studies* 63 (1991).

2.4. 'The Good' as a Theoretical Term

To solve this problem, I suggest, the consequentialiser needs to adopt a certain semantic thesis, about the meaning of the term 'the good' in consequentialism. They need to understand 'the good' as a *theoretical term*.

By a theoretical term I mean a term which is introduced by a theory, in this case consequentialism, as a name for something whose existence the theory postulates.¹⁹ As an analogy, consider so-called 'phlogiston theory'. In the eighteenth century, chemists theorised that all combustible bodies contained a substance that was released in combustion, and they introduced 'phlogiston' as a name for the substance so postulated. As we all know, the theory turned out to be false. It was refuted by the discovery that in fact nothing is released by combustion; there is no such thing as phlogiston. But the theory was immune to refutation in another way: it could not have been refuted by showing that something other than phlogiston was released in combustion. Since 'phlogiston' is a theoretical term introduced by phlogiston theory, it names anything, if there is anything, which has all the properties attributed to phlogiston by this theory, viz. being released by combustion. Were there to exist anything with these properties, then that thing, by definition, would be phlogiston. (For now I ignore the possibility of multiple realisations. If it had turned out that more than one substance was given off in combustion, then phlogiston theory might also have been refuted.)

My suggestion, then, is that the consequentialiser treats 'the good' as a theoretical term introduced by consequentialism, in the same way that 'phlogiston' was introduced by phlogiston theory. That is, 'the good' names the thing, if there is one, which has the properties attributed to the good by consequentialism. The good is thus, by definition, the thing which ought to be maximised.

The analogy is perhaps not perfect. In the case of phlogiston theory, the term 'phlogiston' was not used, as we might say, 'outside' the the-

19. Cf. David Lewis, 'How to Define Theoretical Terms', *The Journal of Philosophy* 67, no. 13 (1970): 427–446.

ory. Proponents of the theory *coined* this term for their own theoretical purposes. However, the term 'the good' is used outside consequentialism. Perhaps this so, but it's not a problem for my proposal. We can think of other cases where an ordinary word is co-opted and, by stipulation, given a purely theoretical meaning. Think of 'string' and string theory, or, closer to home, 'utility' and utility theory. Thus it should be emphasised that in treating 'the good' as a theoretical term, in the way I've suggested, we are giving it a stipulated meaning, which need not be the same as its meaning is in ordinary language. Some might object that this results in an illegitimate definition of consequentialism. A proper definition, they might say, must use the term 'the good' as we ordinarily use it.²⁰ If so, that would be bad news for consequentialisers. But, as I've said, I want to make things as easy for them as I can. Hence I'll allow them to stipulate the meaning of 'the good'.

How does this help the consequentialiser? Recall the moral theory M and its consequentialist counterpart $C\&G$. Let 'the good_G' denote whatever is the good according to G . As we know, an action is right according to M iff it is right according to $C\&G$. And an action is right according to $C\&G$ iff it maximises the good_G. Therefore, an action is right according to M iff it maximises the good_G. That is, M implies that the good_G (and nothing else) ought to be maximised. Given our semantic thesis, M thus implies that the good_G is the good. So the consequentialisation of M is now complete; it has been shown to be a consequentialist theory.

2.5. The Importance of the 'The'

You might worry that our semantic thesis has already made consequentialism a mere tautology (or at any rate, analytically equivalent to one). Consequentialism states that the good ought to be maximised, where 'the good' is defined as the thing which ought to be maximised. If we substitute definiens for definiendum, we thus end up with the trivial looking statement 'The thing which ought to be maximised ought to be maxim-

20. Cf. Mark Schroeder, 'Teleology, Agent-Relative Value, and 'Good'', *Ethics*, no. 117 (2007): 265–295.

ised'. However, while this may seem vacuously true, in fact it isn't. Recall phlogiston theory, which states that phlogiston is released by combustion. Substituting definiens for definiendum as before yields 'The thing which is released by combustion is released by combustion'. This, too, may look vacuously true. But we know it isn't, because we know it isn't true. And we know why: because nothing is released by combustion.

The presence of the definite article 'The' is crucial here. If the statement were 'Anything which ought to be maximised ought to be maximised', then it truly would be trivial; but that's a different, non-equivalent statement. Without getting into controversial details in the philosophy of definite descriptions, we can I think safely make the following two assumptions. First, a sentence of the form 'The F is G' is true iff the definite description 'The F' denotes something which is G. Second, a thing is denoted by 'The F' iff it, and only it, is F. It follows that a sentence of the form 'The F is F' is true iff exactly one thing is F. So the sentence 'The thing which ought to be maximised ought to be maximised' is true iff exactly one thing ought to be maximised.

Thus there remains a way for a moral theory to resist consequentialisation, by implying *nothing* ought to be maximised. (There's another way, by implying more than one thing ought to be maximised.) As I'll argue, some familiar moral theories do resist consequentialisation in this way.

3. The Limits of Consequentialisation

3.1. Formal Model

I begin by defining a simple formal model. The model will take entities of two sorts as primitives: possible worlds and agents. By possible worlds I mean, in the usual way, complete ways the world might be. And by agents I mean creatures capable actions apt for moral evaluation. Let W and I be the sets of respectively all possible *worlds* and all *agents*. For simplicity, assume W is finite. Say a *choice situation* is an ordered pair (i, A) with $i \in I$, $A \subseteq W$, and $A \neq \emptyset$. The worlds in A are the alternat-

ives between which the agent i must choose, the outcomes of the various actions which she might perform in this situation. Let set of all choice situations be $C = I \times \wp(W) \setminus \emptyset$.²¹

What I earlier called the 'deontic output' of a moral theory will then be represented by a mapping R from C into W such that $R(i, A) \subseteq A$. The worlds in $R(i, A)$ are the alternatives which are *right* (permitted) in the situation (i, A) ; the worlds in $A \setminus R(i, A)$ are the alternatives which are *wrong* (forbidden).

The good (if it exists) will be represented by a total preorder \preceq on W , where the position of a world in this preorder represents its goodness.²² That is, $w \preceq w'$ iff w' is at least as good as w . Say that a total preorder \preceq on W *ought to be maximised* iff for all $(i, A) \in C$, $R(i, A) = \max_{\preceq} A$, that is, iff the right alternatives in any given situation are precisely those which, of the alternatives, are greatest or maximal under \preceq .²³

Now consequentialism may be defined in terms of this model as the conjunction of the following to conditions:

Existence (E) some \preceq ought to be maximised;

Uniqueness (U) if \preceq_1 ought to be maximised and \preceq_2 ought to be maximised, then $\preceq_1 = \preceq_2$.

E and U are jointly equivalent to the condition that *exactly one* \preceq ought to be maximised. As can easily be shown, however, U is implied by E; if some \preceq ought to be maximised, then necessarily it is unique in this respect. So we needn't worry about uniqueness. The only question is whether *some* \preceq ought to be maximised.

21. $\wp(X)$ is the power set of X .

22. A binary relation \preceq is a total preorder iff it satisfies three conditions:

1. $x \preceq x$ (reflexivity),
2. if $x \preceq y$ and $y \preceq z$, then $x \preceq z$, (transitivity)
3. if $x \not\preceq y$ then $y \preceq x$ (completeness).

23. $\max_{\preceq} X$ is the set of *maximal elements* in X , i.e. $\max_{\preceq} X = \{x \in X : \forall y \in X (y \preceq x)\}$.

There are in fact some R relative to which *no* \leq ought to be maximised. These represent moral theories which cannot be consequentialised. Which ones are these? To give a clearer idea of this, it will be helpful to state three further conditions (for simplicity I'll leave quantification implicit):

Agent Neutrality (AN) $R(i, A) = R(i', A)$.

No Moral Dilemmas (NMD) $R(i, A) \neq \emptyset$.

Dominance (D) Suppose $\{w, w'\} \subseteq A \cap A'$, $w \in R(i, A)$, and $w' \notin R(i, A)$. Then $w' \notin R(i, A')$.

E is equivalent to the conjunction of AN, NMD, and D (see Appendix for proof). Thus we may distinguish three sorts of moral theory which can't be consequentialised, corresponding to the three conditions. I'll consider each of these in turn.

3.2. Agent Neutrality

Utilitarianism has well-documented difficulties with people's rights. As anyone who's ever been near an ethics class will tell you, it's not hard to imagine situations in which achieving the greatest total happiness would require violating some poor individual's rights, e.g. by denying them a fair trial, or by exploding them with dynamite. The consequentialist who rejects utilitarianism may attempt to overcome such difficulties by accounting for violations of rights in their theory of the good. They may say that worlds in which rights are violated are worse than worlds in which rights are not violated, at least other things being equal. This would be to adopt what Robert Nozick calls a 'utilitarianism of rights', wherein 'violations of rights (to be minimised) merely would replace the total happiness as the relevant end state in the utilitarian structure.'²⁴

However, as Nozick points out, this approach fails to capture what many feel is a crucial aspect of the moral significance of rights, viz. that

rights act as 'side constraints' against our pursuit of the good. They are such that you ought not to violate a right, even in circumstances where your doing so would result in fewer violations overall.²⁵ Suppose that unless you kill one innocent person, thereby violating one right, someone else will kill two innocent people, thereby violating two. The view of rights as side-constraints says that you shouldn't kill the person, yet Nozick's utilitarianism of rights says you should, since this would minimise the total number of rights violations overall.

This side-constraint rights theory is an example of what are sometimes called 'agent-relative', or 'agent-centred', theories.²⁶ As usually defined, a theory is agent-relative iff it gives different aims to different agents; otherwise it's agent-neutral. Thus, for example, the side-constraint rights theory gives me the aim of seeing to it that *I* don't violate any rights, but gives you the aim of seeing to it that *you* don't violate rights. For another example, some moral theories say that agents have a special obligation to promote the wellbeing of their own family and friends, over and above the obligation they have to promote the wellbeing of people generally.²⁷ In contrast, utilitarianism is agent-neutral, because it gives everyone the same aim: maximise total happiness.

Unsurprisingly, if R represents an agent-relative theory then it violates the condition AN. Suppose R represents the side-constraint rights theory. And let the worlds w_m and w_y be such that in w_m you violate some right and I violate none, but in w_m I violate some right and you violate none. (Surely there are such worlds.) Since the theory says I ought to choose w_m over w_y , and you ought to choose w_y over w_m , we have $R(\text{me}, \{w_m, w_y\}) = \{w_m\}$ and $R(\text{you}, \{w_m, w_y\}) = \{w_y\}$. And this plainly violates AN, because by hypothesis $w_m \neq w_y$. So this theory resists consequentialisation, as do all agent-relative theories. These theories imply that nothing ought to be maximised.

This result might be thought to reveal a certain limitation of our formal

25. Ibid., 29.

26. See e.g. Parfit, *Reasons and Persons*, 27.

27. See e.g. Michael Smith, 'Immodest Consequentialism and Character', *Utilitas* 13, no. 2 (2001): 173–94.

24. Robert Nozick, *Anarchy, State, and Utopia* (Blackwell, 1974), 28.

model. A popular strategy for consequentialising agent-relative theories involves devising agent-relative theories of the good, theories according to which what's good relative to one person need not be good relative to another. But the model doesn't allow the good to be agent-relative in this way. It seems to me the most natural way to expand the model so as to overcome this limitation would be to represent the good by a total preorder of *centred worlds*, rather than worlds, where a centred world is a pair (w, i) consisting of a world w and an agent i . The position of (w, i) in the preorder would then represent the goodness of w relative to i .

Thus let W^* be the set of all centred worlds, i.e. $W^* = W \times I$. Say that a total preorder \preceq^* on W^* ought to be maximised iff for all $(i, A) \in C$, $R(i, A) = \{w \in A : \forall w' \in A((w', i) \preceq^* (w, i))\}$. Then define consequentialism* as the conjunction of the following two claims:

Existence* (E*) some \preceq^* ought to be maximised;

Uniqueness* (U*) if \preceq_1^* ought to be maximised and \preceq_2^* ought to be maximised, then $\preceq_1^* = \preceq_2^*$.

This achieves part of what we were after. Clearly E* does not imply AN, and so E* is compatible with agent-relative theories. However, there's a problem: E* implies the negation of U*; if one \preceq^* ought to be maximised, then many ought to be maximised. Thus consequentialism* is inconsistent. (See Appendix for proof.) Perhaps this shouldn't be surprising. According to an agent-relative theory, we might say, there really is no such thing as *the* good. There's not just one thing which ought to be maximised, but many. There's the good-relative-to-me, the good-relative-to-you, the good-relative-to-that-other-guy, and so on.

3.3. No Moral Dilemmas

By a *moral dilemma* I mean a situation in which a person cannot avoid acting wrongly. In our formal model, a moral dilemma is represented by a situation (i, A) such that $R(i, A) = \emptyset$, so the set of wrong alternatives is $A \setminus R(i, A) = A$, i.e. every alternative is wrong. Moral dilemmas have

been widely discussed by philosophers, some of whom believe they are possible. Moreover, the possibility of moral dilemmas seems to be entailed by familiar moral theories which might be called *absolutist theories*, since they incorporate absolute prohibitions. Consider, for example, a theory which holds that violations of rights are absolutely morally forbidden; it is always wrong in any possible situation to violate a right. Suppose, further, that the catalogue of rights endorsed by this theory is such that sometimes a person cannot help but violate at least one right. Then this theory cannot be represented by an R which satisfies NMD, and so it cannot be consequentialised.

Again, this might be thought to show a limitation of our formal model. In our model, the good is represented by a *total* preorder \preceq . That \preceq is total means that, for all w and w' , either $w \preceq w'$ or $w' \preceq w$. In effect, this rules out *incommensurability*, i.e. pairs of worlds such that neither is better than the other, nor are they equally good. Suppose, then, we were to relax this requirement and allow the good to be represented by a merely partial preorder \preceq , i.e. a preorder which is not total. Then for some A , we'd have $\max_{\preceq} A = \emptyset$, thereby enabling the possibility of a moral dilemma within consequentialism.

However, this approach to consequentialising moral dilemmas seems *ad hoc*. It seems arbitrary to say that, in a choice between incommensurable alternatives, both alternatives are wrong, as opposed to both being right, or both being neither. The following seem natural things for the consequentialist to say:

1. an alternative is right if it's at least as good as all alternatives;
2. an alternative is wrong if it's worse than some alternative.

Given the good is represented by a total preorder, these are exhaustive. But if the preorder is not total, then it might be the case that an alternative is not at least as good as every alternative *and* that it's not worse than any alternative. What should the consequentialist say in this case? The approach suggested above would have the consequentialist say the alternative is wrong in this case. But why not instead say it's right? Or

why not say it's neither wrong nor right? So the approach fails, because it rests on this arbitrary decision.

3.4. Dominance

Condition D may be the least intuitive of the three conditions. It requires the following. Suppose that in a given choice situation, two worlds w and w' are among the alternatives. And suppose in this situation, w is right and w' wrong. Then w *dominates* w' in the following sense: w' cannot be right in *any* situation where w is an alternative; the presence of w is always sufficient to make w' wrong.

To find a moral theory which violates this requirement, we may look to theories which relax the maximising element in consequentialism. To some it has seemed that accepting nothing less than the best, as consequentialists do, is too demanding. They believe it's sometimes okay to do something which is worse than the best, provided it's not too much worse. (How much is too much? Don't ask me.) Theories which are less demanding in this way are sometimes called *satisficing* theories, the idea being that 'satisficing' is a more tolerant version of maximising.²⁸

It will be useful to discuss a particular (made up) satisficing theory. Consider then what I'll 'satisficing utilitarianism', defined as follows: an alternative w is right if and only if the total wellbeing in w is at least great as 0.8 times the total wellbeing in every alternative. Suppose there are two choice situations as shown in the following table.

| Situation 1 | | Situation 2 | |
|-------------|-----------|-------------|-----------|
| Worlds | Wellbeing | Worlds | Wellbeing |
| w_1 | 100 | w_1 | 100 |
| w_2 | 80 | w_2 | 80 |
| w_3 | 120 | w_4 | 60 |

In each situation there are three alternatives, two of which— w_1 and w_2 —are available in both. The difference is that w_3 , which is available in the

28. See e.g. Michael Slote, *Common Sense Morality and Consequentialism* (Routledge, 1985).

first situation, is replaced by w_4 in the second. The numbers in the 'Well-being' columns represent the total wellbeing in each of the worlds. Suppose R represents satisficing utilitarianism. Then we have:

$$\begin{aligned} R(i, \{w_1, w_2, w_3\}) &= \{w_1, w_3\} \\ R(i, \{w_1, w_2, w_4\}) &= \{w_1, w_2\}. \end{aligned}$$

But then R cannot satisfy D. If it did, then since w_2 is wrong in the first situation, where w_1 is present, w_2 would be dominated by w_1 , and so it couldn't be right in the second situation, where w_1 is also present. Thus satisficing utilitarianism cannot be consequentialised.

According to satisficing theories, then, there is no such thing as the good. This might seem a puzzling conclusion. Think of the way we would normally describe the difference between satisficing utilitarianism and regular utilitarianism. These two theories, we would say, share a common theory of the good: they both say that the goodness of a world is given by the total wellbeing it contains. Their difference is over how we should 'respond' to the good: roughly, whereas regular utilitarianism says we must maximise the good, satisficing utilitarianism says it's okay if we merely satisfice it. However, our conclusion entails quite a different picture. The two theories *disagree* about the good: whereas regular utilitarianism says there is such a thing as the good, satisficing utilitarianism says there is no such thing.

But the puzzle is easily resolved if we bear in mind the definition of 'the good' we've been assuming. On this definition, 'the good' names the thing, if there is one, which ought to be maximised. Obviously, when proponents of satisficing theories say something like 'It is sometimes permissible not to maximise the good', they cannot mean 'the good' in *this* sense; that would be incoherent. They must mean it in some other sense.

4. Conclusion

So I conclude that the project of consequentialising all moral theories cannot succeed. I'll close by drawing out one moral of this conclusion,

related to something Dreier says.

Dreier's motivation for consequentialising is that he wants to overcome a certain 'stigma' which he says afflicts defenders of 'common sense morality' when they try to deny consequentialism.²⁹ To deny consequentialism, he says, they must claim that we are sometimes required to do less good than we might, but that claim has a 'paradoxical air'.³⁰ So defenders of common sense morality, who deny consequentialism, are stigmatised as having a seemingly paradoxical position. (Note, the semantic thesis adopted above provides an nice explanation of the air of paradox. If the good is, by definition, the thing which ought to be maximised, then to say that a person ought not to maximise the good is to say that they ought not to maximise the thing they ought to maximise, and that's obviously paradoxical.)

Dreier thinks the way to avoid the stigma is to avoid denying consequentialism. If we consequentialise common sense morality, then defenders of common sense morality need not deny consequentialism. If I'm right, however, this way of avoiding the stigma doesn't work; some elements of common sense morality—i.e. agent-centredness, moral dilemmas, satisficing—prevent its consequentialisation. But there's a better way to avoid the stigma. Non-consequentialists need not accept the paradoxical claim that sometimes we ought not to maximise the good. Instead they can claim that there is no such thing as the good.

Appendix

Proposition 1. *R satisfies E iff R satisfies AN, NMD, and D.*

Proof. The left-to-right implication is obvious. We prove only the converse. Suppose R satisfies AN, NMD, and D. Let i be any element of I . And define \preceq by:

$$w \preceq w' \text{ iff } w' \in R(i, A).$$

29. Dreier, 'Structures of Normative Theories', 24.

30. Ibid., 24.

We need to show that \preceq is a total preorder (i.e. reflexive, transitive, and complete), and that $R(i', A) = \max_{\preceq} A$.

Reflexivity and completeness follow from NMD. To prove transitivity, assume for reductio: $w \preceq w'$, $w' \preceq w''$, and $w \not\preceq w''$. Since $w \not\preceq w''$, the definition of \preceq implies $w'' \notin R(i, \{w, w''\})$, and then NMD implies $R(i, \{w, w''\}) = \{w\}$. It follows by D that $w'' \notin R(i, \{w, w', w''\})$. But since $w' \preceq w''$, $w'' \in R(i, \{w', w''\})$, and so by D $w' \notin R(i, \{w, w', w''\})$. And since $w \preceq w'$, it follows by parallel reasoning that $w \notin R(i, \{w, w', w''\})$. But by definition $R(i, \{w, w', w''\}) \subseteq \{w, w', w''\}$. So $R(i, \{w, w', w''\}) = \emptyset$, and therefore R doesn't satisfy NMD, violating our assumption.

Now we prove $R(i', A) = \max_{\preceq} A$. Since R satisfies AN, it will be sufficient to prove $R(i, A) = \max_{\preceq} A$. Assume first $w \in r(i, A)$, it follows from D and NMD that, for all $w' \in A$, $w \in R(i, \{w, w'\})$. So, by the definition of \preceq , $w \in \max_{\preceq} A$. Next assume $w \notin R(i, A)$. If $w \notin A$ then $w \notin \max_{\preceq} A$. If $w \in A$, then it follows from NMD and D that there exists some $w' \in A$ such that $w \neq w'$ and $w \notin R(i, \{w, w'\})$, in which case $w \notin \max_{\preceq} A$. Therefore $R(i, A) = \max_{\preceq} A$. \square

Proposition 2. *R satisfies E* only if R doesn't satisfy U*.*

Proof. Without loss of generality, assume $I = \{1, 2\}$. Suppose \preceq^* ought to be maximised. Then define \preceq_1^* and \preceq_2^* as follows:

- if $i = i'$, $(w, i) \preceq^* (w', i')$ iff $(w, i) \preceq_1^* (w', i')$ iff $(w, i) \preceq_2^* (w', i')$;
- if $i \neq i' = 1$, $(w, i) \preceq_1^* (w', i')$ and $(w, i) \not\preceq_2^* (w', i')$.
- if $i \neq i' = 2$, $(w, i) \not\preceq_1^* (w', i')$ and $(w, i) \preceq_2^* (w', i')$.

Is easily seen that, since \preceq^* ought to be maximised, \preceq_1^* and \preceq_2^* also ought to be maximised. But by construction $\preceq_1^* \neq \preceq_2^*$. \square

References

Anderson, Elizabeth. 'Reasons, Attitudes, and Values: Replies to Sturgeon and Piper'. *Ethics* 106 (1996): 538–554.

- Brink, David. 'Utilitarian Morality and the Personal Point of View'. *The Journal of Philosophy* 83, no. 8 (1986): 417–438.
- Brown, Campbell. 'Two Kinds of Holism about Values'. *The Philosophical Quarterly* (2007).
- Bykvist, Krister. 'Normative Supervenience and Consequentialism'. *Utilitas* (2003): 27–49.
- Dreier, James. 'Structures of Normative Theories'. *The Monist* 76 (1993): 22–40.
- Feldman, Fred. *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge University Press, 1997.
- Lewis, David. 'How to Define Theoretical Terms'. *The Journal of Philosophy* 67, no. 13 (1970): 427–446.
- . *On the Plurality of Worlds*. Blackwell, 1986.
- Louise, Jennie. 'Relativity of Value and the Consequentialist Umbrella'. *The Philosophical Quarterly* 54 (2004): 518–536.
- McNaughton, David and Piers Rawling. 'Agent-Relativity and the Doing-Happening Distinction'. *Philosophical Studies* 63 (1991).
- Moore, G. E. *Principia Ethica*. Cambridge University Press, 1903.
- Nozick, Robert. *Anarchy, State, and Utopia*. Blackwell, 1974.
- Parfit, Derek. *Reasons and Persons*. Oxford University Press, 1984.
- Pettit, Philip and Geoffrey Brennan. 'Restrictive Consequentialism'. *Australian Journal of Philosophy* 64 (1986): 438–455.
- Portmore, Douglas W. 'Consequentializing Moral Theories'. *Pacific Philosophical Quarterly*, no. 88 (2007): 39–73.
- Railton, Peter. 'Alienation, Consequentialism, and Morality'. *Philosophy and Public Affairs* (1984): 134–71.
- Rawls, John. *A Theory of Justice*. Oxford, 1971.
- Schroeder, Mark. 'Teleology, Agent-Relative Value, and 'Good''. *Ethics*, no. 117 (2007): 265–295.
- Slote, Michael. *Common Sense Morality and Consequentialism*. Routledge, 1985.
- Smart, J. J. C. 'Extreme and Restricted Utilitarianism'. *Philosophical Quarterly* 6 (1956): 344–54.
- Smith, Michael. 'Immodest Consequentialism and Character'. *Utilitas* 13, no. 2 (2001): 173–94.