Title
"The Asymmetry Argument"

Permalink
https://escholarship.org/uc/item/3d40m6nq

Author
Peterson, Martin

Publication Date
2008-08-08

THE ASYMMETRY ARGUMENT

MARTIN PETERSON

ABSTRACT. According to some moral philosophers, every moral theory can be consequentialised. To consequentialise a moral theory means to account for moral phenomena usually thought of in non-consequentialist terms, such as rights, duties, and virtues, in consequentialist ways. This paper offers a new defence of this well-known thesis, based on a probabilistic account of moral obligations. The paper also offers a new answer as to why consequentialising moral theories is important: This yields crucial methodological insights about how to pursue ethical inquires.

1. INTRODUCTION

Some moral philosophers believe that every non-consequentialist moral theory can be *consequentialised*. To consequentialise a moral theory means to account for moral phenomena usually thought of in non-consequentialist terms, such as rights, duties, and virtues, in consequentialist ways. In a recent paper, Douglas Portmore proposes the following general recipe for consequentialising non-consequentialist theories:

> Take whatever considerations that the non-consequentialist 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 non-consequentialist theory yields. (Portmore 2007:39)

To see how consequentialising works in practice, consider the Kantian claim that lying is always wrong. If you face a choice between lying and telling the truth, the consequentialist could mimic the advice offered by Kantians by assigning a high utility to telling the truth and a low utility to lying. Put

in Dreier's (1993:23) words, 'we ... take the features of an action that the theory considers to be relevant, and build them into the consequences.'[1]

If true, the claim that every moral theory can be consequentialised reveals an important asymmetry between consequentialism and other theories. Consequentialists can account for the moral importance of phenomena that are usually thought of in non-consequentialist terms, such as rights, duties, and virtues, whereas the opposite is false of non-consequentialist theories. Rights, duty or virtue-based theories cannot account for the fundamental moral importance of consequences.[2] Because of this asymmetry, it seems it would be preferable to become a consequentialist – indeed, it would be virtually impossible not to be a consequentialist. Call this the asymmetry argument.

The present paper seeks to defend two claims about the asymmetry argument. The first is about its truth. I argue that every moral theory can indeed be consequentialised. However, previous accounts of how to consequentialise moral theories, such a Portmore's, have either produced ordinal rankings of acts, which I argue is unsatisfactory, or cardinal rankings based on technical assumptions that many non-consequentialists find implausible. I therefore develop what I think is a more sophisticated account, based on a well-known representation theorem proved by Luce (1959/2005). I argue that the new account avoids the flaws attributed to the previous ones.

The second claim defended herein is about the relevance of the asymmetry argument. Even if every moral theory can be consequentialised, it is far from obvious that consequentialists should welcome this insight. For example, Brown (2004) argues that this threatens to make consequentialism vacuous, since no particular conclusion then follows from consequentialism about what ought to do be done in any given situation. Or, differently put, what is the point of accepting a moral theory that has virtually no normative implications? I argue that the value of the consequentialiser's claim does not lie in its practical implications, but rather in the methodological insights it gives us about how to pursue ethical inquires.

-----

[1] A number of scholars, including Broome (1991), Oddie and Milne (1991), Carlsson (1996), and Brown (2004), have defended similar views.

[2] How do we show that not every version of consequentialism can be conceived of as a version of some non-consequentialist moral theory? I believe this is easy. An important feature of many versions of consequentialism is that *every* kind of bad consequence can be weighted against some sufficiently large good consequence. Even murdering a friend might be right, if this prevents the death of many others. But standard interpretations of major non-consequentialist theories, such as rights, duty and virtue ethics, imply that murdering a friend is always wrong, no matter the consequences. Therefore, consequentialism can never be shown to be a version of some non-consequentialist theory.

2. Four interpretations

How should we understand the claim that non-consequentialist theories can be consequentialised? It seems that this idea can be interpreted in at least four different ways, which are obtained by combining two distinctions. The first is between weak and strong interpretations, and the second between general and restricted interpretations.

According to *strong* interpretations, every moral theory is deontically equivalent (in a sense to be explained later) to some version of consequentialism. This is because all moral theories – including duty ethics, rights-based theories, and virtue ethics – can be represented in some utility function as claims about consequences. If correct, this shows that the theoretical divide between, say, duty ethics and utilitarianism is no greater than the divide between hedonistic utilitarianism and preferentialism. People advocating rival moral theories just make slightly different claims about how to evaluate consequences.

According to *weak* interpretations, there is a genuine difference between consequentialism and non-consequentialism that gives rise to an asymmetry in explanatory power. Non-consequentialist theories can be consequentialised in the sense that consequentialists can accommodate and explain intuitions about rights, duties, and virtues, but the opposite is false of rival theories. For example, rule utilitarians typically claim that intuitions about rights can be justified and explained by the fact that all things considered, the long-term consequences of respecting certain rights tend to be favourable. However, this is not to say that people really *have* rights in the absolute sense proposed by Locke and Nozick. On a weak interpretation, consequentialism merely provides an alternative justification of a number of widespread moral intuitions, but the theory itself is essentially incompatible with the non-consequentialist theories usually taken to justify those intuitions.

Restricted interpretations hold that the possibility of consequentialising a moral theory reveals an asymmetry between consequentialism and *some* of its rivals, whereas *general* interpretations claim that this asymmetry holds between consequentialism and *all* of its rivals. Many scholars defend restricted interpretations. For example, Sen (1982, 2004) argue that consequentialist theories can recognise moral rights. On Sen's view a violation of a right is an evil that ought to be counted as *intrinsically* bad when evaluating a state of affairs. His point is not that the violation of a right makes people unhappy or frustrates someone's preference. The violation of a right is bad in itself, and this badness can be accounted for in consequentialist terms. Sen's point is that the traditional utilitarian way of accounting for

consequences is too narrow. Rights also matter, intrinsically, when evaluating a state of affairs. Clearly, this account of rights is meant to support to a strong, but restricted, interpretation of the asymmetry argument. Sen *construes* rights in purely consequentialist terms, by making a normative claim about how to evaluate rights-violations.

Naturally, it is also possible to account for duties in consequentialist terms by using the line of thought appealed to by Sen in his justification of rights. In this version of duty ethics, duties have intrinsic value. The duty not to lie and the duty not to commit suicide can thus be accounted for in consequentialist terms. If you tell a lie, a consequence of this act is that a certain amount of intrinsic value in the resulting state of affairs is lost. This holds true even if no one is ever made aware of the fact that you lied.[3]

Consequentialists have also attempted to account for the moral importance of virtues. For instance, Driver (1996, 2001a, 2001b) argues that, 'On my theory the value of all these [character] traits resides in their tendency to produce good consequences'.[5] Therefore, on Driver's view, to adopt certain character traits will lead to good consequences and, consequently, this is a good thing to do. Since Driver construes virtues in consequentialist terms, she is in effect defending a strong interpretation. The correlation between virtues and good consequences is not supposed to be a mere empirical coincidence. The correlation is, according to Driver, an essential one. A virtue consists in bringing about good consequences. As pointed out by Bradley (2005), this is by no means an uncommon view among virtue ethicists, not even among scholars advocating 'classic' non-consequentialist versions of the theory. For example, Foot (1978) famously claimed that, 'virtues are in general beneficial characteristics, and indeed ones that a human being needs to

---

[3]Hare (2000:8.4) famously argues that Kantian duty ethics and consequentialism 'does not stand at opposite poles of moral philosophy' and that 'Kant ... *could* have been a utilitarian, though he *was* not'. Hare's main point is that the notion of impartiality inherent in the categorical imperative supports utilitarianism. On Hare's reading of the categorical imperative, 'I shall be able to will only such maxims as do the best, all in all, impartiality, for all those affected by my action. *And this, again, is utilitarianism*'.[4] That said, there is an important difference between Hare's defence of duty ethics and the defence of outlined here: Hare does not assign intrinsic value to individual duties. He merely claims that the overall structure of Kant's theory, and in particular the notion of impartiality derived from the notion of universalisability, can be accounted for in consequentialist terms. According to Hare, Kant's famous claim that we have a duty not to commit suicide is simply false, and the same goes for many similar duties. Hare argues that his version of utilitarianism can account for the theoretical spirit of Kant's theory, but not for all its practical implications. Because of these circumstances, Hare's position cannot be classified as belonging to any of the four versions of the argument outlined above.

[5]Driver (1996:122-3)

have, for his own sake and that of his fellows'.[6] Arguably, this also rings true to the consequentialist. Thus, it is a mistake to think that consequentialism and virtue ethics are fundamentally inconsistent moral views.

All this said about restricted interpretations of the asymmetry argument, it should be noted that Brown (2004), and Portmore (2007) are mainly concerned with general and strong interpretation. On their view, every moral theory, or at least every 'remotely plausible' theory, can be consequentialised, in the sense that it is deontically equivalent to some version of consequentialism. In what follows I shall focus on the strong and general interpretation, mainly because it seems to be the most challenging and interesting version.

3. The strong and general interpretation

According to the strong and general interpretation of the asymmetry argument *every moral theory is deontically equivalent to some version of consequentialism.* In order to assess this claim, we need to clarify at least three issues: What is a 'moral theory'; what does it mean to say that two theories are 'deontically equivalent'; and what counts as a 'consequentialist' moral theory?

A moral theory can be conceived of as a device that attaches deontic predicates to acts. By a deontic predicate I mean predicates such as 'right', 'wrong', 'permissible', 'forbidden', 'obligatory', and 'supererogatory', and so on. This account of moral theories tallies well with the familiar utilitarian claim that it is 'permissible' to perform an act just in case no other act produces more wellbeing. Moreover, the proposed account can account for the Kantian thought that lying is always 'forbidden', as well as for similar claims made by virtually any other moral theory.

If moral theories are conceived of in this way, it follows that the moral value of every act can be represented numerically in at least an ordinal ranking. Here is an example. When Brutus decides to murder Caesar, he faced a choice among five alternatives (see below). Murdering Caesar was morally forbidden, which we represent by assigning a low number, say 0, to this act. Kindly asking him to resign voluntarily would have been supererogatory, which we representant by assigning a higher number to this act, 2. Putting Caesar in prison falls somewhere between murdering him and asking him to resign; we represent this by assigning the number 1 to this act. Note that the numbers merely carry ordinal information – nothing can be concluded about how *much* better it would have been to put Caesar in prison rather than to kill him.

---

[6]Foot (1978:3)

Act	Deontic property	Ordinal ranking
1. Murder Caesar	Forbidden	0
2. Force Caesar to resign without killing him	Permissible	1
3. Put Caesar in prison	Permissible	1
4. Kindly ask Caesar to resign voluntarily	Supererogatory	2
5. Do nothing	Forbidden	0

Two theories are *deontically equivalent* if and only if both theories produce the same (ordinal) ranking no matter which set of acts they are applied to. Note that two deontically equivalent theories may attach different deontic predicates to a given set of acts, as long as the ranking is the same. For example, theory M_1 may divide acts into 'permissible' and 'impermissible' ones, whereas M_2 divides them into 'right' and 'wrong' ones.

However, the mere fact that every moral theory can be associated with numbers does not prove that every theory can be consequentialised. It remains to show that the numbers are somehow determined by the *consequences* of each act. Fortunately, this additional requirement is not as far-fetched as it might first appear to be. If a duty ethicist promises to return a book to the library, but instead sets off for a holiday in the Caribbean, a consequence of this decision seems to be that the promise to return the book will remain unfulfilled. So perhaps it is not completely bonkers to claim that virtually every moral phenomena may count as a consequence?

In his widely circulated manuscript *Climbing the Mountain* (2006), Parfit defines consequentialism as the view that, 'the rightness of acts depends only on facts about how it would be best for things to go.'[7] This is a very general characterisation. To start with, one might wonder if this means that consequentialism is entirely concerned with future events, and thus ignores things that happened in the past?[8] If so, a temporal criterion would provide an easy way of distinguishing between consequentialism and non-consequentialism, because non-consequentialists frequently assign moral weight to past events, such as promises or legal contracts signed a long time ago. However, as correctly pointed out by Parfit, it would be incorrect to think that consequentialism is essentially future-oriented:

> Consequentialists can believe that the goodness of some outcomes depends in part on facts about the past. On two such views, for example, it would be better if benefits went to people who had earlier been worse off, and it would be bad if people were punished for crimes that they did not commit. (Parfit 2006:233)

---

[7] Parfit (2006:252)

[8] This view seems to be defended by e.g. Lars Bergström in his (1966).

An alternative way of outlining the boundaries of consequentialism is to distinguish between agent-neutral and agent-relative theories. For instance, Sen claims that 'consequentialism is ... strongly associated with a characterisation of morality that would *not* permit evaluator relativity in the evaluation of outcomes'.[9] That a moral theory is agent-relative means that the moral value of a state of affairs depends on the perspective from which the evaluation is made. In analogy with this, a moral view is agent-neutral if and only if the moral value of a state of affairs does not depend on the perspective from which the evaluation is made.[10]

Hedonistic utilitarianism is an example of an agent-neutral view. It holds that the moral value of a state of affairs is the arithmetic sum total of all individuals' happiness in that state. Happiness is assumed to be an empirical property, so the arithmetic sum of happiness will, of course, be the same no matter who carries out the computation. This guarantees that hedonistic utilitarianism is agent-neutral.

Unsurprisingly, agent-relative theories come in different flavours. However, the general idea has been rigorously explained by Bykvist (1996) by means of the following example. Suppose that either (i) I benefit my child and you benefit your child, or (ii) I benefit your child and you benefit my child. In cases like this, agent-relative theories entail, by means of definition, that the first alternative is better than the second. This is because one state is better than another *relative to* an agent; that is, it does not make sense to say that one state is better than another *simpliciter*. A well-known example of an agent-relative view is Aristotelian virtue ethics. In Book IX of the Nichomacean Ethics, Aristotle explicitly claims that the virtuous person cares more about friends than strangers. Hence, if you face a choice between saving two friends or three strangers it may very well hold true that *you* ought to save your two friends, because they are your friends, whereas someone else facing the same choice ought to save the three strangers.

However, it has been convincingly argued that some versions of consequentialism can take agent-relative values into account; see e.g. Broome (1991), Skorupski (1995), Bykvist (1996), and Portmore (2001, 2003). Consider, for example, the transplant case introduced by Foot (1967). A doctor faces a choice between two options. The first is to kill one patient and transplant her heart, liver and kidneys to four needing patients, who will die unless they get the organs. The second option is to let the four needing patients die. What should the doctor do? It seems that it is possible to

---

[9]Sen in Scheffler 1991, p 213, via Bykvist.

[10]Nagel and Parfit argue that the distinction should be spelled out in terms of the agent's *aims* or *reasons*, rather than *values*; in what follows this controversy will be ignored.

fill in the details of this example such that the consequences of killing one in order to save four would be far better than the consequences of letting the needing patients die. This spells problem for consequentialism, because almost no one is prepared to accept that doctors should kill some patients in order to save others. However, by allowing for agent-relative values, this counter-intuitive conclusion can be avoided. The key point is to observe that a doctor should refrain from killing his patient because that would involve a killing by *him*. Hence, the consequences of that option are not optimal when evaluated from *his* perspective. Of course, things might turn out differently if the evaluation is made from the point of view of an external observer. According to agent-relative versions of consequentialism, agents ought to perform acts that are optimal from *their* perspective.

In addition to the two points raised above, critics have emphasised two additional features of non-consequentialism that they argue consequentialists cannot account for, viz. moral dilemmas and supererogation.[11] By definition, a moral dilemma is a situation in which every act is morally wrong, or forbidden. No matter what you do you will do something you ought not to do. Most consequentialists find it hard to see how moral dilemmas could ever arise, because they believe every set of acts must comprise at least one act whose consequences are at least as good as those of all alternatives. However, note that this reasoning relies heavily on the assumption that the ranking of consequences is non-cyclical and complete. If the ranking is cyclical all acts would be wrong, since for every act there would exist some act that produces better consequences. Furthermore, a moral dilemma could arise if some consequences are incomparable, i.e. if it is false that one consequence is better than the other and false that they are equally good. In such a case, the ranking would be incomplete, and it would thus be false that every set of acts must comprise at least one act whose consequences are at least as good as those of all alternatives. So on the account outlined here, a moral dilemma arises if a moral theory attaches the predicate 'impermissible' (or 'wrong') to all elements in a set of alternative acts. This can happen if the ranking of consequences is either cyclical or incomplete. Hence, the proposed account of consequentialising is compatible with the existence of moral dilemmas.

So what about supererogation? By definition, an act is supererogatory just in case it goes above and beyond what is required. From a consequentialist point of view, it is natural to link supererogation to satisficing consequentialism. Some consequentialists believe that one is not always morally required to maximise. Sometimes one merely has to do whatever produces sufficiently good consequences – doing what has best consequences

---

[11]See Brown (2004) and Portmore (2007) for excellent overviews of the literature

is supererogatory. It seems that this idea nicely captures a consequentialist notion of supererogatory acts. Whenever an act goes beyond and above what is morally required, the consequentialist theory attaches the predicate 'supererogatory' to the act in question.

The preliminary conclusion is that consequentialists can account for many of the features that non-consequentialists wish to emphasise. However, it remains to show that this broad understanding of the term 'consequentialism' really captures the central ideas we usually associate with consequentialism. Unlike the authors discussed above, I do not think this is case.

4. The need for a cardinal ranking

Consequentialism is attractive mainly because of its conceptual clarity and precision, not because of its normative recommendations. As pointed out above, consequentialism has virtually no practical implications in itself. Consequentialism is compatible with every possible view about what one ought to do in a particular situation. However, in order to take advantage of the conceptual merits of consequentialism we need more than just an ordinal ranking of acts. Contrary to what is sometimes claimed, I believe the methodological virtues come alive once the consequentialist's ranking is made cardinal.[12] It would be a mistake to think that an ordinal ranking is suficient for taking full advantage of the fact that all moral theories can be consequentialised. For brevity, let 'cardinal consequentialism' refer to consequentialist theories that acknowledge the need for cardinal measurement, and let 'ordinal consequentialism' refer to views that reject this requirement.

By definition, cardinal consequentialists maintain that numbers assigned to acts (and consequences) reflect distances in utility, and sometimes even ratios. It is not merely the case that 2 is more than 1, and 1 more than 0. It even holds true that the distance between 2 and 1 is the same as that between 1 and 0; and if utility is measured on a ratio scale 2 is twice as much as 1.

Now, the conceptual advantages of cardinal consequentialism fall into three categories.

First of all, cardinal consequentialism can – unlike ordinal consequentialism – make sense of the distinction between act and rule consequentialism. Some consequentialists maintain that an act should be performed because its actual consequences are good, whereas others maintain that it should be performed because the average consequences of this type of act in the long run tend to be good. Arguably, if the thesis that every moral theory can be consequentialised is to be of philosophical interest, the consequentialist

---

[12]Brown (2004) and Portmore (2007) argue that consequentialisers merely need an ordinal ranking.

theory one ends up with must be sensitive to this distinction. A consequentialiser who cannot account for the difference between act and rule consequentialism has not succeeded to deliver a theory that deserves the label 'consequentialism'. However, note that only cardinal consequentialism can account for this distinction. Rule consequentialism presupposes that one is able to calculate averages (or at least sum up the utility of different consequences into a sum total) and this requires that we measure utility on a cardinal scale.

The second conceptual virtue that speaks in favour of cardinal consequentialism is that it can be successfully applied to cases in which the morally relevant features of a decision are partially known. Suppose, for instance, that you know the probabilities of a set of consequences. Then, in order to distinguish between views holding that an act is right just in case it maximises expected utility, and views holding that one ought to maximise actual utility, we again need cardinal consequentialism. The principle of maximising expected utility can only be applied if utility is measured on a cardinal scale.

The third argument is related to sequential decision making. Sometimes it makes sense to divide a decision into a sequence of sperate decisions to be taken at different points in time; however, in order to chose a sequence of decisions that maximise the sum total of good consequences, value has to be measured on a cardinal scale. Otherwise it is impossible to tell whether, for example, two acts with pretty good consequences together produce more good consequences than a single act with really good consequences.

For the reasons outlined above, I will in the following focus on cardinal consequentialism. The thesis I wish to examine is thus: Is it true that every moral theory is deontically equivalent to some version of cardinal consequentialism? To the best of my knowledge, Oddie and Milne (1991) are the only authors who have defended this claim. They argue that every moral theory M can be *represented* by some other moral theory M^*, and that M^* is certain to be a version of cardinal consequentialism. More precisely put, they argue that no matter which moral theory one subscribes to, it is always possible to assign cardinal numbers to acts in a way that conforms to the consequentialist dictum that an act belongs to the highest deontic category (i.e. is 'right') if and only if its consequences are at least as good as those of all alternative acts.

Let us take a closer look at Oddie and Milne's argument. They start from the assumption that every moral theory is concerned with a finite number of deontic categories. The purpose of formulating a moral theory is, in their

view, to sort all alternative acts into exactly one of these deontic categories.[13] The deontic ordering produced by the moral theory M is represented by the binary relation $\geq_M$. In line with what we have said above, two moral theories M and M^* are *deontically equivalent* just in case $\geq_M$ and $\geq_{M*}$ order the alternative acts $\{a_1, \ldots, a_n\}$ in the same way.

According to Oddie and Milne, a moral theory is a version of consequentialism just in case: (i) there is a function that assigns value to states of affairs, (ii) the deontic features of an act are determined solely by the value of the act, and (iii) the value of an act is determined by a function V that takes the values and probabilities of the possible consequences of the act as its argument. Oddie and Milne impose three structural conditions on the function V, which entail that the value of an act equals its objective expected utility. Now, the main theorem proved by Oddie and Milne states that for every M, if the expected utility axioms hold for all consequences, then there exists a deontically equivalent theory M^* that fulfils conditions (i) – (iii).

Unfortunately, a major problem with Oddie and Milne's representation theorem is that many non-consequentialists would not accept the premises on which it is based. Oddie and Milne *stipulate* that the value of every consequence can be represented by some real number.[14] From a non-consequentialist point of view, this is an unacceptable assumption. To start with, some non-consequentialists believe that consequences might be incomparable. For instance, an agent facing a genuine moral dilemma, in which he will do something wrong whatever he does, would say that such alternatives cannot be morally compared. It is neither the case that one consequence is better than the other, nor are they equally good. Hence, since the existence of moral dilemmas is excluded by a real-valued representation of consequences, some non-consequentialists have good reasons for rejecting Oddie and Milne's approach.

An additional problem with assigning real numbers to consequences is that this presupposes the Archimedean property of the reals.[15] If one consequence is assigned a high number and another consequence a low number, it can be immediately concluded that the first consequence is a finite number

---

[13]In a comment on the paper by Oddie and Milne, Carlsson (1995:30) argues that this view of a morality is over-simplified, 'since a deontic ordering dos not itself induce an ordering of [the agent's] alternatives'. I think this complaint is mistaken. Arguably, one could expand the set of deontic categories such that each element in the ordering of alternatives corresponds to exactly one deontic category. Cf. Oddie and Milne (1991:52).

[14]Oddie and Milne (1991:56).

[15]The Archimedean property can be formulated as follows: For all x,y, if $x > y$ then there exist some n such that $n \cdot y > x$.

of units better than the latter. However, many non-consequentialists would reject such claims. For instance, defending his rights-based theory, Nozick would claim that the consequence in which a right is violated is *infinitely* worse than the consequence in which no right is violated. Therefore, the value of violating a right, rather that respecting it, cannot be represented as a distance between real numbers.

The weakness of Oddie and Milne's cardinal approach can be further clarified by considering the von Neumann-Morgenstern axiomatisation of the expected utility principle. (Oddie and Milne make no direct reference to the vNM axioms, since they propose a slight different axiomatisations, but because the vNM axioms are all necessary consequences of the expected utility principle they are nevertheless committed to these axioms.) Imagine an agent who is facing a choice between a set of alternative acts $\{f, g, h, \ldots\}$. Assume that you are an external observer, observing the agent's behaviour. Occasionally, there is some uncertainty with respect to which act will be performed by the agent (as viewed from the perspective of the external observer). Perhaps the external observer is sitting on an ethics committee advising others what to do. By assumption, the external observer knows all there is to know about the acts.

Naturally, defending his favourite moral theory, the external observer agrees that it is better that an act deemed to belong to the highest deontic category is chosen with a high probability, rather than with low probability. The external observer is even prepared to express preferences over probabilistic mixtures of alternatives. Hence, the observer may claim that, say, the lottery in which 'f is performed with probability p and g with probability $1 - $ p' is better than the lottery in which 'g is performed with probability q and h with probability $1 - $ q'.[16] What structural conditions, if any, should these preferences over probabilistic mixtures of acts satisfy?

Enter the vNM axioms. In order to state these purely structural conditions, let $F = \{f, g, h, \ldots\}$, and let G be the set of functions $p : F \to [0, 1]$ such that $\sum_{f_i} p(f_i) = 1$. We assume that for every number $\alpha \in (0, 1)$ and every f and g in G, the composite act $\alpha \cdot f + (1 - \alpha) \cdot g$ is also in G. Hence, $G = \{f, g, h, \ldots\}$ is the set of all possible (binary) probabilistic mixtures between the elements of F. Let the relation $\succ$ on G denote a comparative moral evaluation between a pair of its elements, i.e. $f \succ g$ means that f is morally

---

[16]In order to link these preferences to things that matter for the observer, one may imagine some ingenious story in which the observer is an unselfish agent concerned with performing morally good acts; perhaps the agent takes advice from the observer, thereby giving the observer some probabilistic influence over what act will be performed by the agent.

better than g. In this framework, the vNM axioms can be formulated as follows.[17]

> (i) The relation $\succ$ is complete and transitive in G.
> (ii) If $f \succ g$, then $\alpha \cdot f + (1 - \alpha) \cdot h \succ \alpha \cdot g + (1 - \alpha) \cdot h$.
> (iii) If $f \succ g \succ h$, then there exist $\alpha, \beta \in (0, 1)$ such that
> $\alpha \cdot f + (1 - \alpha) \cdot h \succ g \succ \beta \cdot f + (1 - \beta) \cdot h$.

Condition (i) is an ordering condition, which implies that all acts have to be fully comparable from a moral point of view. This assumption is discussed in more detail below. Condition (ii) is the infamous independence condition, which has been extensively criticised by decision theorists.[18] Condition (iii) is an Archimedean condition. It claims that it will always be possible to construct a lottery between something good and bad, which is strictly better than something in between, given that the probability for the good consequence (act) is high enough.

Taken together, conditions (i) – (iii) imply that the moral value of each act f, g, h,... in G can be represented by a utility function that is unique up to a positive linear transformation. This means that the utility of each alternative act can be measured on an interval scale. (Note that utilities are assigned both to probabilistic acts as well as to normal, non-probabilistic acts in G.) Formally put, the representation and uniqueness theorem states that there exists a function u from G to [0,1] such that:

> (1) If $f \succ g$, then $u(f) \succ u(g)$, and
> (2) $u(\alpha \cdot f + (1 - \alpha) \cdot g) = \alpha \cdot u(f) + (1 - \alpha) \cdot u(g)$, and
> (3) for every other function u' satisfying (1) and (2), there are numbers $c > 0$ and d such that $u' = c \cdot u + d$.

The proof of this result is reproduced in many text books on decision theory.[19] However, non-consequentialists have strong reasons to reject conditions (i) and (iii). The reason for this is exactly the same as the reason why Oddie and Milne cannot use real numbers for representing the value of consequences. To start with, consider the objection that the completeness assumption [Condition (i)] is incompatible with the existence of moral dilemmas. Arguably, a moral dilemma is best conceived of as a case in which

---

[17]von Neumann and Morgenstern expressed their axioms in a slightly different way; see their (1947: 24-27). The formulations given here, which are more attractive from a technical point of view, can be found in e.g. Kreps (1988: 43-4) and Schmidt (1998: 4-6).

[18]See e.g. Kreps (1988) for an overview.

[19]See e.g. Resnik (1993).

the relation $\succ$ is *not* a complete order, i.e. a case in which some alternatives are incomparable. For instance, if you cannot avoid violating either your brother's rights or your sister's, it is neither better, nor worse, nor equally as good to prefer one option over the other.

The Archimedean condition [Condition (iii)] is also unacceptable from a non-consequentialist point of view. Both Kant and Nozick would agree that an act deemed wrong by their theories, such as lying or stealing, in *infinitely* worse than right acts. Nozick explicitly claims that 'Imposing how slight a probability of a harm that violates someone's rights also violates his rights'.[20] However, the Archimedean condition contradicts this intuition. If f and g are two fairly good acts such that f is slightly better than g, and h is some morally horrible act, such as a murder, then the Archimedean condition guarantees the existence of a lottery in which f is performed with a high probability and h with a very low probability, which is strictly better than the performance of g with certainty. The kind of probabilistic trade-off inherent in the Archimedean condition is simply not acceptable according to several versions of non-consequentialism.

What should one make of these objections? Some non-consequentialists obviously believe in moral dilemmas, so if the consequentialiser cannot model this idea he seems to be doomed. The same is true of the non-Archimedean feature of many prominent non-consequentialist theories. However, I believe neither objection poses any genuine threat. In fact, not even all consequentialists can accept the vNM axioms. The consequentialist endorsing Rawls's maximin condition must, for instance, reject the Archimedean condition. Of course, it would be incorrect to conclude that such non-Archimedean theories are not versions of consequentialism. It would also be incorrect to conclude that such theories cannot be mathematically represented by utility functions. All that follows is that not every moral theory can be represented by a von Neumann-Morgenstern function. In the literature of measurement theory, several alternative measurement systems have been proposed. A non-Archimedean (lexicographic) expected value function was even briefly considered by von Neumann and Morgenstern (1947: 631), and full-blown non-Archimedean analyses have been detailed by others. For example, by dropping the Archimedean condition and replacing it with a strengthened version of the independence condition, Hausner (1954) obtained a vector representation of value.[21] In a similar vein, Carlsson (2006) has shown that the completeness assumption can be dropped, given that we are prepared to accept a vector representation.

---

[20]Nozick (1974: 74)

[21]See also Skala (1975) and Schmidt (1998)

Naturally, in order to show that every moral theory can be consequentialised, it is sufficient that there is *some* way in which every moral theory can be represented by a utility function. It is not necessary that there is a *single* way in which all theories can be represented. Despite this, the next section will nevertheless propose a single way in which to represent all moral theories. A unified representation is attractive mainly because this is simpler than using different representations for different theories.

5. A unified approach

In contrast to previous arguments for the claim that every moral theory is deontically equivalent to some version of cardinal consequentialism, the argument to be developed here does not require full comparability, nor does it presuppose that moral value is Archimedean. Moreover, the new argument employs a more restricted ontology, in which acts, consequences, and value orderings are defined in terms of states of affairs.

A state of affairs consists of all facts that obtain at a certain point in time. It is plausible to assume that all moral theories can be formulated as moral claims about states of affairs. For instance, act consequentialism assigns moral value to the future states actually brought about by an act. Rule consequentialism, on the other hand, evaluates a state brought about by an act in relation to how similar states typically, in the long run, promote wellbeing. In a similar vein, duty ethics is the view that it is morally wrong to bring about a certain *kind* of state. The set of forbidden states constitutes the kinds of states that it is our duty to avoid bringing about. For example, if it is a duty not to murder people, then it is morally wrong to bring about states in which someone is murdered. In analogous ways, if it is a duty not to break promises, then it is morally wrong to bring about states in which promises are broken. In both examples, the moral value of an act depends on facts that obtain at different points in time, viz. an earlier fact in which the victim is alive or a promise is given, as well as a fact occurring later, in which the victim is murdered or the promise is broken. By reasoning in analogous ways, other non-consequentialist views can also be conceived of as evaluative claims about states of affairs.[22]

_____

[22]According to the natural rights tradition advocated by Locke and Nozick, rights exist somehow independently of us. Locke and Nozick are hence committed to the view that every act that brings about a certain type of state (a states in which a right is violated) is wrong. A right is thus nothing more and nothing less than the wrongness of bringing about a certain undesirable state. For the virtue ethicist, an act is morally right if it brings about the same state, and is triggered by similar (mental) states, as what would have been the case had the act been performed by a fully virtuous agent. In order to spell out this ethical theory entirely in terms of states, it thus seems that a rather large number of states

An act can be conceived of as an ordered pair $\langle s, c \rangle$, where s is an initial state of affairs and c is the state brought about by the act, i.e. its 'consequence'. Consider, for instance, the assassination of Kennedy by Oswald. The state s is the state that obtained just before Oswald pulled the trigger, in which Kennedy was sitting in his black limousine in Dallas. The consequence c is the state that obtained shortly afterwards, in which Kennedy was dead. In this context, the term 'consequence' does, of course, not merely refer to the amount of happiness or preference satisfaction in the world. If a duty ethicist promises a dying friend to remember her after her death but then ignores that promise once she is dead, the state in which the promise has been broken is a consequence of the act of ignoring to honour the memory of the friend. This holds true even if all well-being levels remain the same. In the case of promise-breaking, s is thus a state that includes both the fact that a promise has been made as well as the fact that the promise is still respected, whereas c is a state that includes the fact that a promise has been broken.

It now remains to explain how a cardinal measure over the set of acts is to be obtained. Briefly put, the measurement proposed here is based on a distinction between two notions of acts, viz. *real-world acts* and *descriptions* of real-world acts. When we say that acts should be conceived of as functions we refer to descriptions of acts. Hence, when the moral philosopher sets up a thought experiment in which killing one in order to save four has better consequences than refraining from actively killing anyone, he is merely describing the abstract contours of some particular acts. Although statements about the moral status of descriptions of acts can perhaps be made with absolute certainty, statements about the moral status of particular, real-world acts will always be more or less uncertain. There is a huge difference between *stipulating* that some act in a thought experiment would lead to better consequences overall than all alternatives, and saying that a particular *real-world* act would lead to better consequences overall. The second statement will always be riddled with some uncertainty. For example, if I pull the trigger of my gun the bullet will probably kill you, and therefore

have to be considered. Firstly, one has to determine the initial state obtaining before the act was performed, as well as the state actually brought about by the agent. However, the moral value of this transformation of one state into another is calculated by comparing it with how closely it corresponds to a similar hypothetical transformation of one state into another performed by a fully virtuous agent. Note that this way of formulating virtue ethics makes it very similar to e.g. hedonistic utilitarianism. The latter maintain that one should evaluate transformations of one state into another solely by adding the amounts of happiness created in the new state; the virtue ethicists adopts a slightly more complex evaluative mechanism, which involves a comparison with an ideal agent.

it is highly probable that the act in question is morally wrong. However, it is *possible* that the bullet would instead have accidentally killed the person standing on your right, who might one day have become a cruel dictator, in which case the act might have been morally right. In similar ways, there may always be some doubt about the actual motive triggering one's act, or some doubt about whether a promise was actually made ten minutes ago.

In what follows, the possibility of moral doubt will play a crucial role. It is helpful to make this assumption as explicit as possible: *All* claims about the moral ranking of acts are susceptible to error. Moral uncertainty can be modelled by ascribing probabilities to which act has the highest deontic status. In what follows, we shall ascribe probabilities to statements such as, 'act f has the highest deontic status in the set of alternatives B according to the moral theory M'. This probabilistic notion roughly corresponds to Oddie and Milne's deontic betterness-relation $\geq_M$. If you accept a moral theory M according to which you have a duty never to break a promise, and you are fairly certain that you promised the librarian to return *Naming and Necessity* today, the probability might be, say, .99 that you are obliged to return the book instead of keeping it for another week. However, there is a small probability that you are not obligated to return the book, since it is not entirely certain that you actually promised the librarian to return it. These probabilities can be interpreted either subjectively or objectively. If interpreted subjectively, they tell us what the agent subscribing to M *believes* about the moral status of some act. If interpreted objectively, the most plausible interpretation is, arguably, to maintain that the probabilities reveal information about what an ideal, fully rational agent *ought* to believe about the moral status of some act. The moral theory M attaches deontic predicates to abstract descriptions of acts, which invariably prompts the agent accepting M to ascribe deontic probabilities to the corresponding real-world acts. Thus, on this version of the conseqentialiser's thesis, every version of non-consequentialism will be equivalent to some version of consequentialism on the most concrete level, i.e. when it comes to making moral evaluations of real-world acts.

This probabilistic framework is compatible with non-Archimedean moral theories. This is because the agent can never be entirely sure what act is best. One merely ascribes probabilities to qualitative comparisons between acts, and one can never be sure that one alternative is deontically better than another. The agent is thus never justified in claiming with full certainty that, say, not murdering a person is certain to be infinitely better than doing so. In the present framework, the non-Archimedean conviction that murdering people is infinitely worse than not doing so will simply be represented by a subjective probability close to one that one ought not to murder. The probabilistic approach is moreover compatible with the existence of moral

dilemmas. No matter whether the agent faces a genuine moral dilemma, or a situation that just appears to be a dilemma, it holds true for each dilemmatic alternative that there is some probability (greater than zero) that it has a higher deontic status than the other alternative according to theory M. The insight that neither option is permissible does not affect the fact that a choice has to be made. Of course, an agent facing a moral dilemma may feel horrible 'on the inside' compared to an agent facing a set of non-dilemmatic options, but when it comes to choosing what to do, there is no fundamental difference.

There is an interesting link between the deontic probability of an act (the probability that is has a higher deontic status than some alternative act relative to a moral theory M) and its utility. Let $p\,(f{\succ}_M B)$ denote a probability of p that act f has the highest deontic status in B according to the moral theory M. Let $p\,(A{\succ}_M B)$ denote the probability of the following conditional: If A is a subset of B, then the probability is p given M that the top element in B is also an element of A.

We will now take advantage of the famous choice axiom proposed by Luce (1959/2005). This axiom holds that if A is a subset of B, then the probability that f would be chosen from B equals the probability that f would be chosen from A multiplied by the probability that the chosen alternative in B is also an element of A. In symbols,

AXIOM: Let $A \subset B$. Then $p(x \succ B) = p(x \succ A) \cdot p(A \succ B)$.

To grasp what kind of assumption is at stake in this axiom, it is helpful to consider an example from the decision theoretical context in which it was originally proposed. Suppose a lady visiting a posh restaurant is about to choose a wine from a list containing two red and two white wines. The axiom states that it should not matter if the woman divides her choice into two stages, that is, first choose between red and white wine, and then between the wines in the chosen subset, or choose directly which of the four wines to order. Hence, if the woman is indifferent to the choice between red and white wine in general, as well as to the choice between the two red wines and the two white ones at hand, the probability that a particular bottle will be chosen is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. In the present context, the probabilities, of course, refer to how certain it is that one option is morally better than another.

The choice axiom implies the existence of a cardinal utility scale.

THEOREM 1 [Luce 1959/2005] If the choice axiom and the axioms of the probability calculus hold for B and all its subsets, then there exists a positive real-valued function u on B such that for every $A \subset B$, $p\,(f{\succ}_M A) = \frac{u(f)}{\sum_{g \in A} u(g)}$,

and for every other function u' satisfying this condition there is a constant k such that $u = k \cdot u'$.

The proof is straightforward and follows Luce (1959/2005:23-24): Since it was assumed that $p(A \succ B) \neq 0$, the choice axiom implies that:

$$(1) \qquad p(x \succ A) = \frac{p(x \succ B)}{p(A \succ B)}$$

In order to prove the existence part, saying that a utility function u exists, let $u(x) = k \cdot p(x \succ B)$, where $k > 0$. Then, since the elements of A are mutually exclusive, the probability axioms guarantee the truth of the following equation.

$$(2) \qquad p(x \succ A) = \frac{k \cdot p(x \succ B)}{\sum\limits_{y \in A} k \cdot p(y \succ B)} = \frac{u(x)}{\sum\limits_{y \in A} u(y)}$$

In order to prove the uniqueness part, suppose that u' is another function defined as above. Then, for every $x \in B$, the following holds.

$$(3) \qquad u(x) = k \cdot p(x \succ B) = \frac{k \cdot u'(x)}{\sum\limits_{y \in B} u'(y)}$$

By letting $k' = k / \sum\limits_{y \in B} u'(y)$, it immediately follows that $u(x) = k'u'(x)$.

Q.E.D.

6. Conclusion

It seems that it is always possible to measure how well things go by assigning utilities on a cardinal scale to alternative acts, provided there is some uncertainty about the deontic status of each act. No substantial conclusion follows from this about what ought to be done in a particular situation.

So why is this result important? If every moral theory can be mimicked by some version of consequentialism, it seems that little or nothing has been gained – this theoretical manoeuvre seems to be of little help. Brown (2004: 3) even suggests that, 'If every non-consequentialist theory may be ... mimicked [by some version of consequentialism], then consequentialism is vacuous'.

That said, I believe the significance of the thesis that every moral theory can be consequentialised is not that it offers any action guidance. Its significance rather stems from the fact that it helps us gain a better understanding

of moral theories, especially of theories that have previously been thought of as non-consequentialist views. This point is worth pushing a bit further. Imagine a future in which moral theorists engage in ranking alternative acts as described above. Naturally, utilitarians and duty ethicists will continue to disagree over the moral value of breaking a promise. However, this disagreement would be no deeper than disagreements within utilitarianism itself. Some utilitarians believe that happiness matters, whereas others focus on objective lists of intrinsically valuable items. Both versions of consequentialism can be rendered very precise by applying the conceptual machinery available within the consequentialist framework.

In the foregoing sections I have emphasised three such conceptual tools available to the consequentialist, which are currently unavailable to traditional non-consequentialists, viz. 1) the distinction between act and rule based versions moral theories, 2) the distinction between the actual consequence and expected consequences of an act, and 3) the importance of sequential decision making. For example, if we consequentialise duty ethics it would make sense to distinguish between versions of duty ethics according to which one ought to act such that as many duties as possible are *actually* fulfilled, and versions according to which it is the *expected* duty-fulfillment that matters. Traditional accounts of duty ethics cannot make much sense of this useful distinction. This indicates that something important is actually gained by consequentialising a non-consequentialist theory such as duty ethics.

Therefore, instead of trying to render theories usually thought of as non-consequentialist more precise by using the rather blunt conceptual tools available within those frameworks, the insight that every moral theory can be consequentialised suggests that more precision can be gained by expanding the consequentialist circle to encompass views that have traditionally been thought of as non-consequentialist.

References

[1] Bradley, B. 2005. 'Virtue Consequentialism',*Utilitas* 17: 282-298.
[2] Brandt, R. 1992. *Morality, Utilitarianism, and Rights*. Cambridge: Cambridge University Press.
[3] Broome, J. 1991. *Weighing Goods*. Oxford: Basil Blackwell.
[4] Brown, C. 2004. 'Consequentialise This', unpublished manuscript.
[5] Bykvist, K. 1996. 'Utilitarian deontologies? On preferemce utilitarianism and agent-relative value',*Theoria*, 62:124-143.
[6] Carlsson, E. 1995. *Consequentialism Reconsidered*. Theory and Decision Library, Kluwer Academic Publishers.
[7] Carlsson, E. 2006. 'Incomparability and Measurement of Value'. In K. McDaniel et al. (ed.), *The Good, The Right, Life, and Death* Aldershot: Ashgate.

[8] Driver, J. 1996. 'Virtue and Human Nature', in R. Crisp (ed.) *How should one Live?*, Oxford.

[9] 2001a. *Uneasy Virtue*. New York: Cambridge University Press.

[10] ed. 2001b. 'Character and Consequentialism'. Special Issue of *Utilitas*, 13:2.

[11] Foot, P. 1967. 'Abortion and the Doctrine of Double Effect', *Oxford Review*, 5: 28-41.

[12] Foot, P. 1978. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Blackwell.

[13] Foot, P. 1983. 'Utilitarianism and the Virtues', *Proceedings and Addresses of the American Philosophical Association*, 57, 2: 273-83.

[14] Hare, R. M. 2000. 'Could Kant have been a utilitarian?', chapter 8 in his *Sorting Out Ethics* Oxford: Oxford University Press.

[15] Kreps, D. 1988. *Notes on the Theory of Choice*, Westview Press.

[16] Locke J. 1993.*Two treatises of government*. London: Dent.

[17] Luce, D. 1959/2005. *Individual Choice Behavior. A Theoretical Analysis*. John Wiley and Sons.

[18] Nozick, R. 1974. *Anarchy, State, and Utopia*. Oxford : Blackwell.

[19] Oddie, G.& Milne P. 1991. 'Act and Value: Expectation and the Representability of Moral Theories', *Theoria* 57: 42-76.

[20] Parfit, D. 2006. *Climbing the Mountain*, manuscript.

[21] Pettit, P. 1997. 'The Consequentialist Perspective' in *Three Methods of Ethics*, by M. Baron, P. Pettit, and M. Slote. Oxford: Blackwell.

[22] Portmore, D. 2001. 'Can an Act-Consequentialist Theory be Agent-Relative?' *American Philosophical Quarterly*, 38: 363-77.

[23] Portmore, D. 2003. 'Position-Relative Consequentialism. Agent-Centered Options, and Supererogation', *Ethics*, 113: 303-32.

[24] Portmore, D. 2007. 'Consequentializing Moral Theories', Pacific Philosophical Quarterly 88: 39-73.

[25] Schmidt, U. 1998. 'Axiomatic Utility Theory under Risk: Non-Archimedean Representations

[26] Skala, H. 1975. *Non-Archimedean Utility Theory*, Springer.

[27] Sen, A. 1982. 'Rights and Agency', *Philosophy and Public Affairs*, 11:3-39.

[28] Sen, A. 2004. 'Elements of a Theory of Human Rights', *Philosophy and Public Affairs*, 32:315-56.

[29] von Neumann, J. & Morgenstern, O. (1947) *Theory of Games and Economic Behavior,* 2nd edition, Princeton University Press.

DEPARTMENT OF HISTORY AND PHILOSOPHY OF SCIENCE, UNIVERSITY OF CAMBRIDGE

E-mail address: `mbp24@cam.ac.uk`