Philosophical Books Vol. 48 No. 4 October 2007 pp. 303–319

MEASURING LIFE'S GOODNESS

DAVID MCCARTHY
The University of Edinburgh

One major strand of ethical theorizing proceeds by supposing that there exist what I will call *individual goodness measures*. These are functions from lotteries over entire world histories to the real numbers which somehow purport to measure how good such lotteries are for individuals. Utilitarianism plainly presupposes that such measures exist when it says that one lottery is better than another just in case it contains a greater sum of individual goodness. And so do various well known ways of departing from utilitarianism.

Given the credentials of this approach to theorizing, it is surprising how little attention philosophers have paid to the question of what individual goodness measures are, or indeed whether they exist. But if we lack good answers to these questions, it is hard to be confident about the approach.

One philosopher who has not neglected such questions is John Broome. His first treatment of them runs through the whole of his book *Weighing Goods* (henceforth *WG*).[1] But in his more recent book *Weighing Lives* (*WL*) he largely withdraws his earlier views and offers an alternative treatment.[2] In this article I try to chart both positions. But I will also say why I prefer a third.

WG and *WL* may well be the most subtle and penetrating works in the utilitarian tradition. But this makes them far from easy to understand. Critical distance is needed, and I often prefer to put their arguments in my own words. The material until the end of section 4 outlines these arguments, and sections 5 onwards take up the critical study and development of an alternative.

1. Interpersonal and Intrapersonal Comparisons

I begin with a single framework for expressing both intrapersonal and interpersonal comparisons. My approach is similar to the one used in *WL*, though my terminology differs. For example, where *WL* uses subjective expected utility theory, I simplify by using objective expected utility theory.

Let i and j be (the same or different) individuals, and L_1 and L_2 be lotteries over entire world histories. The usual way of expressing intrapersonal and interpersonal comparisons is by saying such things as: facing L_1 would be at least as good for i as facing L_2 would be for j. In facing these lotteries, the lives the individuals will lead will depend on how the lotteries turn out. I can therefore paraphrase the expression by saying: (i,L_1) is at least a good a lottery

1. John Broome, *Weighing Goods* (Blackwell, 1991).
2. John Broome, *Weighing Lives* (Oxford University Press, 2004).

Philosophical Books Vol. 48 No. 4

over lives as (j,L_2). I will call the corresponding relation the *better life lottery relation*.

I will assume without question that the better life lottery relation satisfies the axioms of expected utility theory. The only problematic assumption that involves is that the better life lottery relation is *complete*: for any individuals i and j, and any lotteries L_1 and L_2, either (i,L_1) is at least a good a lottery over lives as (j,L_2), or (j,L_2) is at least a good a lottery over lives as (i,L_1). Some will doubt this assumption because they believe that there are incommensurable goods. What to say about individual goodness measures in the face of incompleteness is an important problem, but it is not the problem I am discussing here. I will therefore treat the completeness assumption as a useful simplifying assumption.

A real valued function *f represents* a binary relation R just in case: for all x and y, Rxy if and only if $f(x) \geq f(y)$. Given that the better life lottery relation satisfies the axioms of expected utility theory, there exists a function $u(i,L)$ which represents it such that for any individual i and any lottery $L=[p_1,h_1; p_2,h_2; \ldots ; p_m,h_m]$, where $h_1 \ldots h_m$ are histories and $p_1 \ldots p_m$ are their corresponding probabilities, all summing to one,

(1) $u(i,L)= p_1u(i,h_1) + p_2u(i,h_2) + \ldots + p_mu(i,h_m)$

This function is known as an *expectational* function because the value it assigns to the life lottery equals the value it assigns to the various possible lives that could result, multiplied by their probabilities then added up. Throughout, $u(i,L)$ will always be this function.[3]

2. Weighing Goods I

The *WG* analysis comes in two parts. This section outlines the first part. As I understand it, with universal quantification implicit, the first part of the *WG* analysis tries to fill out the schema

(2) The goodness of the life (j,h_2) rather than (i,h_1) existing is at least as great as the goodness of the life (l,h_4) rather than (k,h_3) existing iff ____.

WG does not attempt to fill out (2) in full generality, and only looks at two special cases. The first essentially involves just one person. Suppose A is the sole member of the population, and suppose that (A,h_2) is a better life than (A,h_1), and that (A,h_4) is a better life than (A,h_3). Consider the following two lotteries. The two bearers of uncertainty, or *states of nature*, heads and tails, are equally likely.

3. For further details, see *WG*, Ch. 5; and Peter Hammond, 'Interpersonal comparisons of utility: why and how they are and should be made', in J. Elster and J. Roemer (eds.), *Interpersonal Comparisons of Well-Being* (Cambridge University Press, 1991), pp. 200–254.

L_1		L_2	
Heads	*Tails*	*Heads*	*Tails*
h_2	h_3	h_1	h_4

Table 1

Which is better, L_1 or L_2? The consideration in favour of L_1 is that in state of nature heads, the life (A,h_2) rather than (A,h_1) exists. The consideration in favour of L_2 is that in state of nature tails, (A,h_4) rather than (A,h_3) exists. This generalizes Broome's example, but his remark still applies: "these two considerations—each a difference in good—have to be weighed against each other" (*WG*, p. 146). Hence

(3) The goodness of the life (A,h_2) rather than (A,h_1) existing is at least as great as the goodness of the life (A,h_4) rather than (A,h_3) existing iff L_1 is at least as good as L_2.

This is only a staging post to the filling out *WG* aims for. The next step is to provide an analysis of when L_1 is at least as good as L_2. *WG* argues for

> *The principle of personal good* Assume a constant population. Then if two lotteries L and M are equally good for each person, they are equally good; and if L is better for at least one person than M, and at least as good for everyone, then L is better than M. (*WG*, p. 165)

This principle entails that L_1 is at least as good as L_2 if and only if L_1 is at least as good for A as L_2. Since $u(i,L)$ represents the better life lottery relation, (1) entails that L_1 is at least as good for A as L_2 iff

$$\tfrac{1}{2}u(A,h_2) + \tfrac{1}{2}u(A,h_3) \geq \tfrac{1}{2}u(A,h_1) + \tfrac{1}{2}u(A,h_4)$$

which is equivalent to

$$u(A,h_2) - u(A,h_1) \geq u(A,h_4) - u(A,h_3).$$

Combining this with (3), we get

(4) The goodness of the life (A,h_2) rather than (A,h_1) existing is at least as great as the goodness of the life (A,h_4) rather than (A,h_3) existing iff $u(A,h_2) - u(A,h_1) \geq u(A,h_4) - u(A,h_3)$.

The next step in the first part of the *WG* analysis is to try to further fill out (2) by looking at cases essentially involving just two people. Suppose A and B are the only two and consider two histories h_5 and h_6, where (A,h_5) is a better life than (A,h_6), and (B,h_6) is a better life than (B,h_5).

Which is better, h_5 or h_6? The consideration in favour of h_5 is that the life (A,h_5) rather than (A,h_6) exists. The consideration in favour of h_6 is that the life

(B,h_6) rather than (B,h_5) exists. This generalizes an example in *WG* (p. 215), but the same idea still applies. These two considerations—again, a difference in good—have to be weighed against each other. Hence

(5) The goodness of the life (A,h_5) rather than (A,h_6) existing is at least as great as the goodness of the life (B,h_6) rather than (B,h_5) existing iff h_5 is at least as good as h_6.

To make progress, we need to say more about when h_5 is at least as good as h_6. A variation on a theorem due to Harsanyi helps.[4] Say that the betterness relation is *impartial* just in case: for any lotteries L_1 and L_2 involving a constant population, if there exists a permutation π of the population such that for every member i of the population, L_1 is exactly as good for i as L_2 is for $\pi(i)$, then L_1 and L_2 are equally good.

Constant population theorem: Assume a constant population of individuals i=1 . . . n. Suppose
(i) the better life lottery relation satisfies the axioms of expected utility,
(ii) the betterness relation satisfies the axioms of expected utility,
(iii) the principle of personal good is true, and
(iv) the betterness relation is impartial.
Then there exists an expectational function $u(i,L)$ which represents the better life lottery relation such that $\Sigma_i\, u(i,L)$ represents the betterness relation.

Cutting a long story short, the only premise for which there is not a good prima facie case is the simplifying assumption that the better life lottery relation is complete. But we have already taken that assumption on board as an idealizing assumption. This version of Harsanyi's theorem follows easily from the original. It differs from the central result of *WG*, the interpersonal addition theorem, by having an explicit way of expressing interpersonal comparisons, and is then able to follow *WL* by being explicit about impartiality.

If the premises of the constant population theorem are correct, h_5 is at least as good as h_6 if and only if

$$u(A,h_5) + u(B,h_5) \geq u(A,h_6) + u(B,h_6)$$

which is equivalent to

$$u(A,h_5) - u(A,h_6) \geq u(B,h_6) - u(B,h_5).$$

Combining this with (5) we get

(6) The goodness of the life (A,h_5) rather than (A,h_6) existing is at least as great as the goodness of the life (B,h_6) rather than (B,h_5) existing iff $u(A,h_5) - u(A,h_6) \geq u(B,h_6) - u(B,h_5)$.

4. See John Harsanyi, 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', *Journal of Political Economy*, 63 (1955), pp. 309–21.

WG does not offer what could be interpreted as a filling out of (2) in the most general case, but it is not hard to generalize the approach *WG* takes.[5] Let h_7, h_8, h_9 and h_{10} be histories each of population size one, and let the people in these histories be *A*, *B*, *C* and *D*, with no assumption made about whether any of these are the same person or not. Suppose (B,h_8) is a better life than (A,h_7), and that (D,h_{10}) is a better life than (C,h_9). Let L_3 be the lottery [h_8 if heads, h_9 if tails], and L_4 be the lottery [h_7 if heads, h_{10} if tails].

Which is better, L_3 or L_4? I lack space to discuss this, but the main idea is to extend the definition of impartiality to capture the idea that the identities of persons do not matter. Using the constant population theorem, one can run a similar argument as before to conclude that

(7) The goodness of the life (B,h_8) rather than (A,h_7) existing is at least as great as the goodness of the life (D,h_{10}) rather than (C,h_9) existing iff $u(B,h_8) - u(A,h_7) \geq u(D,h_{10}) - u(C,h_9)$.

Moreover, given the premises of the constant population theorem, (4), (6) and (7) could be generalized by adding people who are unaffected by the choices between the histories or lotteries in question. So by looking at what *WL* calls weighing goods across the dimensions of states of affairs, lives and people, we have arrived at progressively more general ways of filling out (2).

3. Weighing Goods II

The second part of the *WG* analysis moves from (4) and (6) to a conclusion about individual goodness measures. The argument would work just as well if applied to (7).

Suppose the two lotteries L_1 and L_2 are equally good. By (4),

(8) The goodness of the life (A,h_2) rather than (A,h_1) existing equals the goodness of the life (A,h_4) rather than (A,h_3) existing if and only if $u(A,h_2) - u(A,h_1) = u(A,h_4) - u(A,h_3)$.

Slightly altering his passage to match my terminology, Broome says:

> Since the two differences in good are exactly balanced in determining the overall goodness of the lotteries, it would be very natural to express this fact by saying that the differences are actually the same. That is
>
> $$g(A,h_2) - g(A,h_1) = g(A,h_4) - g(A,h_3)$$
>
> These differences count the same in determining overall goodness. So to deny they are actually the same would be to insist on a distinction between amounts of good and how much those amounts count in determining

5. Compare *WL*, p. 92 and Hammond, 'Interpersonal comparisons of utility'.

overall goodness. And it is natural to think this an empty distinction. (*WG*, p. 147)

The suggestion is that the function $u(i,h)$ is a measure of how good it is that the life (i,h) exists. This conclusion is drawn from a discussion of weighing goods across states of nature. And *WG* takes the same conclusion to be supported by the discussion of weighing goods across lives. For by applying (6) to the case in which the two histories h_5 and h_6 are equally good, the same form of argument will conclude that it is very natural to say that

$$g(A,h_5) - g(A,h_6) = g(B,h_6) - g(B,h_5)$$

And Broome concludes that a "distinction between quantities of good and how much those quantities count now seems very empty" (*WG*, p. 217). The discussion of weighing goods across lives therefore adds to the support for the claim that $u(i,h)$ is a measure of how good it is that the life (i,h) exists.

I will skip the technical details, but the above arguments rely upon the fact that $u(i,L)$ is an expectational function. But any old expectational function which represents the better life lottery relation would have done, so the interim conclusion of the *WG* analysis is this. The class of functions which measure how good it is that lives exist is (a nonempty subset of) the class of expectational functions which represent the better life lottery relation. The qualification about subsets allows for the possibility that we can narrow down the class of functions which measure how good it is that lives exist still further. Section 8 discusses this issue, but for now we can ignore it.

But how does this connect with individual goodness measures? Individual goodness measures take what *WL* calls a personal perspective: they measure how good histories are for individuals. This fits with the way they are related to utilitarianism. To assess how good a history is, utilitarians assess how good it is for the first individual, how good it is for the second, and so on, and then add all these up. But the discussion so far in this section has been about how good it is that a particular life exists. This takes what *WL* calls an external perspective. *WG* identifies these perspectives. It assumes, but does not make explicit, what I will call the *identification thesis*: a measure of how good a life is for the individual who leads it is a measure of how good it is that that life exists.

Combining the identification thesis with the interim conclusion of the *WG* analysis gives

> *The risk neutrality thesis* The class of individual goodness measures is (a nonempty subset of) the class of expectational functions which represent the better life lottery relation.

WG argues for what Broome calls *Bernoulli's hypothesis*: if a lottery L gives i at least as great expectation of individual goodness as a lottery M, then L is at least as good for i as M. The risk neutrality thesis is equivalent to the conjunction of Bernoulli's hypothesis together with the idealizing assumption that the better life lottery relation is complete. But it does not hurt to think of

Bernoulli's hypothesis and the risk neutrality thesis as the same, and I will speak as if Broome is arguing for the latter.

4. Weighing Lives

WL somewhat distances itself from the *WG* analysis. To begin, having made the identification thesis explicit, *WL* denies that it is a conceptual truth. The view that the wellbeing of future people should be discounted is supposed to show this. For that view "plainly does make sense, even if it is wrong" (*WL*, p. 93). But if that view is correct, the identification thesis is false, for two lives can be equally good for the two people who lead them even though one counts more towards general goodness because it exists earlier (*WL*, pp. 93–4). Since temporal discounting at least makes sense, the identification thesis is said not to be a conceptual truth.

But why does *WL* think that *WG* was committed to the identification thesis as a conceptual truth, rather than just a plain ethical truth? I am not certain about this, but I think this points to an important difference between *WG* and *WL*. Where *WG* offers sustained argument towards a single ethical view, *WL* seems sympathetic to the claim that a small number of conflicting views about weighing goods are reasonable. But it may then seem implausible that a claim about meaning could be derived from a single one of those views. Talk of individual goodness is the common property of ethical theorists with different views, and it may seem implausible that the meaning of such talk is to be derived from a single such view, even if it is correct.

If I am wrong about this interpretation, it at least makes sense of other remarks *WL* makes. For *WL* claims that *WG* was mistaken in seeking a precise analysis of the meaning of our quantitative talk about individual goodness. The intuitive meaning we assign to such talk is "vague" and we do not actually have a "precise meaning". We can make this vague meaning precise, but we have a choice about how to do that (p. 90).

I suggest that this plurality of reasonable choices emerges from a plurality of reasonable views about weighing. To illustrate, we can see why *WL* thinks that the *WG* analysis is not entirely wasted.

Return to Table 1, and suppose now that L_1 and L_2 are equally good *for A*. Mimicking the discussion in previous sections, instead of (8) we could conclude with

(9) The goodness for A of the life (A,h_2) rather than (A,h_1) existing equals the goodness for A of the life (A,h_4) rather than (A,h_3) existing if and only if $u(A,h_2) - u(A,h_1) = u(A,h_4) - u(A,h_3)$.

And mimicking the argument that followed, we could conclude that it would be very natural to express the fact that L_1 and L_2 are equally good for A by saying that

$$g(A,h_2) - g(A,h_1) = g(A,h_4) - g(A,h_3).$$

But there would be no mention of general goodness in any of this discussion, only goodness for A, and we would now directly interpret $g(A,h)$ as a measure of how good the life (A,h) is for A. The identification thesis has been bypassed, but we still end up with the risk neutrality thesis.

But other reasonable views about weighing goods would lead to just as natural, but different ways of measuring individual goodness (*WL* 90, 226). So *WL* adopts the risk neutrality thesis as a reasonable but mildly arbitrary definition which goes beyond our ordinary meaning.

5. The Default Thesis

I now begin the critical study of the *WG* and *WL* analyses. This section describes what I will call the default thesis, a well known position which conflicts with the *WG* and *WL* analyses. Based on some brief remarks made in *WG*, section 6 outlines the best response I can think of for *WG* and *WL*. But section 7 points to a difficulty with this response, and uses it to motivate an alternative approach I call the network analysis. That gives us three competing analyses of individual goodness measures, and section 8 extends the contrast by looking at what *WL* says about some issues about population size. The final section 9 draws conclusions about which analysis is best, and ends by endorsing the network analysis.

An obvious necessary condition for a function to be an individual goodness measure is that it represents the better life lottery relation. But it is far from clear that there is any other condition a function has to satisfy to be an individual goodness measure. So the following seems well named. In different terminology, many welfare economists have thought it obviously correct.

> *The default thesis* The class of individual goodness measures is the class of functions which represent the better life lottery relation.

By contrast, the risk neutrality thesis says that the class of individual goodness measures is (a subclass of) the class of expectational functions which represent the better life lottery relation. So how much difference does adding 'expectational' make?

A lot. Skipping technical details,[6] the default thesis says that there is no quantitative structure to individual goodness. The risk neutrality thesis says that there is a lot of quantitative structure. So the risk neutrality thesis makes a stronger claim than the default thesis. Since the default thesis is a popular and apparently quite plausible view, it is fair to ask why Broome thinks it is wrong.

In fact, friends of the default thesis will have two concerns about Broome's views. *WG* sets out to analyze the meaning of our quantitative talk about

6. For these see, e.g., *WG*, Ch. 6, and David McCarthy, 'Utilitarianism and Prioritarianism I', *Economics and Philosophy*, 22 (2006), pp. 335–363.

individual goodness. But friends of the default thesis will ask: what quantitative talk? And ordinary language seems to be on their side. For it is hard to believe that the folk do anything at all like talk quantitatively about how good lives are for people. And when philosophers do so by offering examples which are meant to trigger our allegedly anti-utilitarian intuitions (100 units of individual goodness for *A* and 0 for *B* versus 45 each), the proper reaction should be to ask what on earth is being said.

This points to the first concern. The *WG* analysis starts with examples which can be described in terms of differences, and infers that these differences reflect an underlying quantitative structure. But although the term 'the difference between __ and __' is sometimes used to refer to an arithmetical operation, oftentimes it is not. Since they think that our talk of individual goodness has no real quantitative structure to start with, friends of the default thesis will think that the *WG* inference is question begging and trades on two different senses of that term. And they will think, not with *WL* that it is only mildly indeterminate which quantitative structure talk of individual goodness picks out, but that it picks out no quantitative structure at all. Adopting the risk neutrality thesis as a definition covers up far more arbitrariness than *WL* admits.

The second concern starts by granting for the sake of argument that our talk of general and individual goodness has a lot of quantitative structure, but then claims that it is not obvious why the risk neutrality thesis gives the right account of that structure. For suppose we grant a central part of the *WG* analysis, that any quantitative account of individual goodness has to be able to describe when the facts described on the left hand side of the 'if and only if' in (8) obtain. But (8) says that they obtain if and only if

$$(10)\quad u(A,h_2) - u(A,h_1) = u(A,h_4) - u(A,h_3).$$

The *WG* analysis says that the correct description of when (10) obtains, in terms of an individual goodness measure $g(i,L)$, is

$$(11)\quad g(A,h_2) - g(A,h_1) = g(A,h_4) - g(A,h_3).$$

But now let $g(i,L)$ be any function which represents the better life lottery relation but which, when interpreted as an individual goodness measure, is inconsistent with the risk neutrality thesis. A function f is *increasing* just in case for all x and y, if $x > y$ then $f(x) > f(y)$. One can show that for some increasing but not linear function f, $f(g(i,L)) = u(i,L)$. So (10) is now true if and only if

$$(12)\quad f(g(A,h_2)) - f(g(A,h_1)) = f(g(A,h_4)) - f(g(A,h_3)).$$

This alternative to the risk neutrality thesis can therefore describe in full generality when the allegedly crucial fact obtains. So the *WG* analysis has to claim that there is somehow something right about (11) and wrong about (12) as descriptions of the crucial fact. But what? And how does this connect with the *WL* analysis?

6. The Simplicity Argument

Broome is aware of such concerns, but his reply is brief. What is wrong with (12) is that it introduces an "empty distinction" (*WG*, pp. 147, 217). An empty distinction is a distinction which does no real work. And (12) certainly seems to be a more complicated description of the way we weigh goods than (11), and hence to be needlessly complicated. Where the description in (11) only uses an individual goodness measure and the simple operation of subtraction, (12) uses an additional mathematical transformation, the function *f*. So it is in the spirit of *WG* to see what is right about the risk neutrality thesis as having something to do with simplicity, and what is wrong about the alternatives as having something to do with needless complexity. What I call the simplicity argument goes beyond Broome's words, but tries to develop this idea.

To begin, utilitarianism turns out to presuppose that there exists at least one individual goodness measure and that any individual goodness measure can be turned into any other individual goodness measure by a positive affine transformation, a transformation of the form $x \rightarrow ax + b$ where $a > 0$.[7] And if this presupposition is satisfied, then if it is restricted to constant population problems, the betterness relation according to utilitarianism (for short: the utilitarian betterness relation) is easily seen to be complete.

An increasing transformation is a transformation of the form $x \rightarrow f(x)$ where f is an increasing function. It is easy to show that if some function represents the better life lottery relation, so do all increasing transformations of that function. But the class of positive affine transformations is a tiny subset of the class of increasing transformations. So for the utilitarian presupposition to be satisfied, the class of individual goodness measures has to be tiny in comparison with the class of all functions which represent the better life lottery relation.

Therefore, if the default thesis is true, utilitarianism suffers from presupposition failure. In fact, the failure is severe. For example, if the default thesis is true, the content of utilitarianism is easily seen to be no different from the content of prioritarianism, the view that we should depart from utilitarianism by giving greater priority to the worse off. But prioritarianism is usually seen as one of utilitarianism's main competitors. So the default thesis convicts both utilitarianism and the large fragment of ethical theory which surrounds it of massive error.

There is therefore strong interpretative pressure to reject the default thesis and to vindicate the utilitarian presupposition. But it is hard to see how the default thesis could be false, for the only obvious constraint a function has to satisfy to be an individual goodness measure is to represent the better life lottery relation.

But this is just an instance of a general problem. For it often happens that all the obvious constraints on interpretation are too weak to avoid an

7. See McCarthy, 'Utilitarianism and Prioritarianism I'.

intolerable error theory about a large body of discourse. Compare, for example, Kripke's discussion of whether to interpret 'plus' in terms of adding or quadding.[8] But there are well developed solutions to this problem, and one seems tailor made.

Lewis argues that properties come in degrees of naturalness.[9] And the more natural a property is, the more eligible it is to be interpreted as the referent of a predicate. So suppose two candidate referents satisfy all the apparent constraints on the interpretation of a term, but there is pressure to interpret just one of them as the referent of the term. Then other things being equal, the more natural candidate referent should be interpreted as the referent. And when Lewis applies this idea to the interpretation of mathematical terms like 'plus', he seems to regard simplicity as contributing towards naturalness.

Putting all this together, there is pressure to interpret 'is an individual goodness measure' in terms of the positive affine transformations of some function which represents the better life lottery relation. But there are many different ways of doing this, hence the threat of an intolerable error theory. But if we can show one such way is distinguished in terms of simplicity, Lewis's idea will solve our problem.

And it appears that we can show this. Suppose we interpret an expectational function which represents the better life lottery relation as an individual goodness measure $g(i,L)$. Then for any individual i and any lottery $L = [p_1,h_1; p_2,h_2; \ldots; p_m,h_m]$

$$(13)\quad g(i,L) = p_1 g(i,h_1) + p_2 g(i,h_2) + \ldots + p_m g(i,h_m).$$

This interpretation is appealingly simple. It makes the goodness for i of facing the lottery L just equal to the goodness for i of each life i might have as a result, multiplied by its probability then all added up.

But now suppose we interpret some nonexpectational function which represents the better life lottery relation as an individual goodness measure $g(i,L)$. Then $f(g(i,L)) = u(i,L)$ for some increasing but not linear function f. Since f is increasing it has an inverse f^{-1}, and it is simple algebra to show that

$$(14)\quad g(i,L) = f^{-1}[p_1 f(g(i,h_1)) + p_2 f(g(i,h_2)) + \ldots + p_m f(g(i,h_m))].$$

But if f is increasing but not linear, one can show that f^{-1} and the f's do not cancel out. So (14) cannot be further simplified, and seems more complicated than (13). So the nonexpectational interpretation seems to provide a less simple individual goodness measure than the expectational interpretation. Moreover, any two expectational functions which represent the better life lottery relation are always related by a positive affine transformation. So the risk neutrality thesis vindicates the utilitarian presupposition, and Lewis's views about interpretation make the case for it.

8. Saul Kripke, *Wittgenstein on Rules and Private Language* (Harvard University Press, 1982).
9. See David Lewis, 'New Work For a Theory of Universals', *Australasian Journal of Philosophy*, 61 (1983), pp. 343–77; and 'Putnam's paradox', *Australasian Journal of Philosophy*, 62 (1984), pp. 221–236.

That is the simplicity argument. It goes beyond *WG*, but I think it makes sense of the *WG* analysis. It explains what the naturalness of (11), along with the complaint that (12) introduces an "empty distinction", has to do with meaning. And it avoids relying on the identification thesis. Moreover, it also makes sense of the *WL* analysis. What underlies the simplicity of (11) and the more general (13) are their additive forms. But without going into the technicalities, what makes additive representations possible is the idea that we can, in Broome's metaphor, weigh goods across the corresponding dimension—across states of nature, in the case of (13). While *WG* suggests that all cases where we can weigh goods point to the same conclusion about individual goodness measures, *WL* thinks there is a plurality of reasonable ways of weighing goods which do not all point to the same conclusion. But on any view, there are few dimensions across which it is reasonable to weigh goods (in more technical terms: few dimensions across which the evaluative relation in question is strongly separable, hence additively representable). So *WL* will hold that there are few additive representations of reasonable ethical views. But the pressure to vindicate the utilitarian presupposition persists. So if simplicity plays the role in *WL* I have suggested it plays in *WG*, no wonder *WL* claims that it is indeterminate how to interpret individual goodness measures, but only mildly. The risk neutrality thesis is well motivated, but not mandated.

7. The Network Analysis

The simplicity argument rests on controversial claims about simplicity and its role in constraining interpretation. But I will focus on a more specific difficulty, and use it to motivate an alternative to the *WG* and *WL* analyses.

The risk neutrality thesis entails that some expectational function which represents the better life lottery relation is an individual goodness measure. When we are discussing utilitarianism, it does not matter which expectational function as long as we restrict ourselves to constant population problems, which I will do throughout this section. So we can take a shortcut by assuming that $u(i,L)$ is an individual goodness measure. It then follows that U is the utilitarian betterness relation, where U holds between two lotteries L_1 and L_2 involving a constant population just in case $\Sigma_i u(i,L_1) \geq \Sigma_i u(i,L_2)$. So if the simplicity argument is a good argument, it provides good reasons to accept the risk neutrality thesis. And the risk neutrality thesis logically entails that U is the utilitarian betterness relation. But I will claim that the simplicity argument does not provide good reasons to accept that U is the utilitarian betterness relation. Therefore, it is not a good argument.

The reason for this is that we already have a substantial body of platitudes about utilitarianism and its place in ethical theory which any exercise in radical interpretation must try to vindicate. For example, given the centrality of utilitarianism's place in ethical theory, accepted by friends and foes alike, it is a platitude that utilitarianism expresses important, reasonably well motivated ethical ideas. It is also a platitude that in some important way utilitarianism ignores the separateness of persons and is insensitive to matters

of distribution. Then there are comparative platitudes about, for example, the way utilitarianism is related to views like prioritarianism and egalitarianism. I do not say that these platitudes are especially clear, or that we might not come to revise them. Nevertheless, any proposal which commits us to an account of what the utilitarian betterness relation is must vindicate these platitudes well enough or else earn a failing mark.

However, the simplicity argument ignores these platitudes, and it is highly nonobvious whether the claim that U is the utilitarian betterness relation does vindicate them. For example, when asked what's so special about $u(i,L)$ in comparison with the nonexpectational functions which represent the better life lottery relation, the simplicity argument answers: it's simple. But the fact that $u(i,L)$ is simple does not by itself give us any reason to believe that the claim that U is the betterness relation expresses important, well motivated ethical ideas. Anyone who claims otherwise is fetishizing mathematics, not doing ethics.

Similarly, if the utilitarian betterness relation between histories is represented by $\Sigma_i u(i,h)$, each prioritarian betterness relation between histories is represented by $\Sigma_i w(u(i,h))$ for some increasing and strictly concave function w. It is said that prioritarianism respects the separateness of persons while utilitarianism ignores it.[10] But why does maximizing the sum of something simple ignore it, and the sum of something not so simple respect it? This is opaque. Likewise, why is maximizing the sum of something simple distributively insensitive in any way we should be concerned with, and maximizing the sum of something not so simple distributively sensitive? This too is opaque.

There is little doubt that the simplicity argument does not by itself show that interpreting U as the utilitarian betterness relation vindicates the various platitudes about utilitarianism. It does not even address them. So for all the simplicity argument says, it is an open question whether U deserves to be interpreted as the utilitarian betterness relation. Since the simplicity argument logically entails that U is the utilitarian betterness relation, there is, at best, a gap in it. It may seem unfair that I am criticizing Broome for an argument which is not clearly his. But since neither *WG* nor *WL* ask whether interpreting U as the utilitarian betterness relation vindicates the various platitudes about utilitarianism, there is a gap in them as well.

To fill this gap, we need to think about which candidate betterness relation best vindicates our network of interrelating platitudes about utilitarianism when interpreted as the utilitarian betterness relation. Elsewhere I argue that interpreting U as the utilitarian betterness relation does indeed vindicate these platitudes well enough, and better than any competing interpretation. This argument does not make any assumptions about individual goodness measures other than that they represent the better life lottery relation. And it does not make any assumptions about simplicity and its role in constraining interpretation. However, the claim that U is the utilitarian betterness relation straightforwardly entails the risk neutrality thesis.[11] This approach applies the

10. See Derek Parfit, *Equality or Priority?* (Lindley Lectures, University of Kansas, 1991); and Wlodek Rabinowicz, 'Prioritarianism for Prospects', *Utilitas*, 14 (2002), pp. 2–21.
11. See McCarthy, 'Utilitarianism and Prioritarianism I' for details.

Ramsey-Carnap-Lewis treatment of implicit definition, and I will call it the *network analysis*.[12]

8. Zeroing In

The *WG* and network analyses conclude with the risk neutrality thesis, but disagree on the grounds for that conclusion. This section discovers a similar phenomenon when we look at what the *WL* and network analyses say when the population size is allowed to vary.

Say that a life (k,h) is a *neutral existence* just in case h is exactly as good as $h_{-k,}$ where (i) h_{-k} is some history which has the same population as h except that k does not exist in $h_{-k,}$ and (ii) for every member i of the population of $h_{-k,}$ h is exactly as good for i as h_{-k} (*WL*, p. 141). So a neutral existence is a life whose existence makes the history it is part of neither better nor worse. We can likewise define good and bad existences.

Suppose that according to utilitarianism, (k,h) is a neutral existence. The definition of utilitarianism in the introduction then entails that all individual goodness measures give (k,h) a value of zero. Since utilitarianism already presupposes that any two individual goodness measures are related by a positive affine transformation, it now follows that utilitarianism makes the stronger presupposition that any two individual goodness measures are related by a positive linear transformation, a transformation of the form $x \rightarrow ax$ for some $a > 0$.

Given the risk neutrality thesis, to define the class of individual goodness measures in a way which satisfies this presupposition it is enough to arbitrarily pick some expectational function which represents the better life lottery relation, then let the class of individual goodness measures be the positive linear transformations of that function.

Skipping technicalities, this approach arbitrarily assigns a value of zero to the individual goodness of some life or other. This life could be what is intuitively a very good life, or a very bad life. Once this choice has been made, utilitarianism will have been arbitrarily defined as a theory which among other things entails that lives of that value are neutral existences.

In effect, this is the approach *WL* takes. It adopts a zero only as a convenient normalization (p. 254). However, this approach will be unsatisfactory to someone sympathetic to the network analysis if it is already implicit in ethical discourse that utilitarianism has a view about what a neutral existence is like.

I suggest that it is implicit. Following *WL*, say that a life is *constantly neutral* just in case at each time t during the life, dying at t is as good for the person who is leading it as living the rest of the life. We can likewise define constantly good and constantly bad lives. I suggest that according to utilitarianism, constantly neutral lives are neutral existences. I will only be able to make two brief remarks in defence of this suggestion. But if it is right, it would yield

12. See David Lewis, 'How to Define Theoretical Terms', *Journal of Philosophy*, 67 (1970), pp. 427–446. I borrow the term 'network analysis' from Michael Smith, *The Moral Problem* (Blackwell, 1994).

The zero thesis The individual goodness of constantly neutral lives is zero.

Here is the defence. Suppose instead that utilitarianism follows *WL* in claiming that constantly neutral lives are bad existences (similar remarks apply to the claim that they are good existences). A continuity argument would then establish that some constantly good life is a bad existence. Utilitarianism would then say that it is worse to create such a life, but that once created, it is at all times better to maintain it. So in a perhaps loose sense, utilitarianism becomes dynamically inconsistent, taking different attitudes to the same life at different times.

I suggest that this position coheres poorly with utilitarian ideas for two reasons. The substantive reason is that the kind of value which would have to underlie such a position looks somewhat alien to the utilitarian tradition. The structural reason is that according to both the *WG* and network analyses, the claim that the betterness relation satisfies the axioms of expected utility theory underlies utilitarianism.[13] But it is arguable that a concern with dynamic consistency in turn underlies expected utility theory.[14]

If these conjectures are correct, the claim that constantly neutral lives are neutral existences coheres well with ideas already associated with utilitarianism. Such coherence may not be strong enough to make the zero thesis already implicit in our ethical discourse. But we make theoretical progress by uncovering definitions which cohere with past discourse and which make future theorizing more precise. Given an otherwise arbitrary choice, it would not take especially strong coherence to make the zero thesis the best way of defining the zero.

These conjectures would take a separate article to defend, but they illustrate how by appealing to views we already hold about utilitarianism the network analysis could be applied to defining the zero. As it happens, the zero thesis is the definition *WL* adopts, but only on the grounds of convenience. So now it is the *WL* and network analyses which may agree in conclusion, but disagree on the grounds for that conclusion.

9. Push and Pull

I complained that the *WG* and *WL* analyses ignore our network of platitudes about utilitarianism and its place in ethical theory. Given the definition of utilitarianism, these platitudes are implicitly in part about individual goodness measures. Call all such platitudes the *individual goodness platitudes*. The issue throughout has been what fixes the referent of our quantitative talk of individual goodness, and any account of this which ignores the individual goodness platitudes seems doubtful. But I will end by discussing a response to this criticism.

13. See *WG*, Ch. 10, and McCarthy, 'Utilitarianism and Prioritarianism I'.
14. See Peter Hammond, 'Consequentialist Foundations for Expected Utility', *Theory and Decision*, 25 (1988), pp. 25–78.

To begin I will suppose that conditions are as favourable as possible for the *WG* analysis. So I will assume that *WG* has established beyond all doubt that *U* is the betterness relation, and that all truths about weighing goods point to the risk neutrality thesis. Thus I will imagine that interpreting an expectational function which represents the better life lottery relation as an individual goodness measure provides an overwhelmingly simpler way of organizing the truth about ethics than so interpreting any nonexpectational function. Call these the *simplicity assumptions*. And I will take Lewis's views about reference for granted.

Lewis argued that the referents of our terms are determined by two forces.[15] There is our linguistic practice and intention to correctly describe the world. This pushes our terms towards the world in ways which make what we say and think come out true. But there is also the greater eligibility of the more natural properties to be interpreted as the referents of our terms. This pulls our terms towards the natural properties even if that makes what we say and think come out false.

This general position leaves open the relative importance of these two forces, and the defence of *WG* I am now imagining gives great weight to the pull. As soon as we try to organize our ethical thinking by talking quantitatively about individual goodness, this weight is great enough that the pull of their simplicity automatically makes this talk lock on to the expectational functions. *Contra* the *WL* analysis, the various reasonable but actually incorrect ways of weighing are irrelevant; *contra* the network analysis, so are the individual goodness platitudes.

This is the best defence I can find of the *WG* analysis. But I am afraid I do not believe it. One of Lewis's goals was to refute an immoderate position due to Putnam that truth and ideal long run verification coincide.[16] Putnam's argument works by allowing the referents of our terms to be highly unnatural properties. Lewis argued that this focuses only on the push, and that the pull of natural properties is strong enough to make such unnatural properties ineligible as the referents of our terms. This restores the moderate view that truth is not so easy to come by.

The current defence of *WG* is immoderate in a parallel but opposite way. If the pull of natural properties is so strong that our terms lock on to them no matter how much we get wrong about them, referring to such properties is too easy. Giving greater weight to the push of our intention to correctly describe the world restores the moderate view that reference to the natural properties is not so easy. *Contra* the current defence of *WG*, the individual goodness platitudes are important in fixing the referent of our quantitative talk about individual goodness. So even under conditions which are as favourable as possible for the *WG* analysis, the network analysis still plays a major role.

But what of the *WL* analysis? This is at its most promising in one of two cases. In case 1, not all the truths about weighing goods point in the same direction. In case 2, the truths about ethics are not what our talk of individual

15. See Lewis, 'New Work for a Theory of Universals', and 'Putnam's Paradox'.
16. Hilary Putnam, 'Models and Reality', *Journal of Symbolic Logic*, 45 (1980), pp. 464–482.

goodness is primarily aimed at anyway. The latter case will obtain if as non-cognitivists think there are no such truths, or if it is the class of reasonable views that our talk of individual goodness is aimed at organizing, not the one true view.

But in case 1, the role of the network analysis will be greater than its role under the conditions most favourable to the *WG* analysis. And in case 2, if the point is to organize the way we think about ethics, the network analysis has the advantage that it does not ignore much of this thinking.

Finally, what of *WL*'s claim that *WG* is mistaken in seeking a precise analysis of the meaning of our quantitative talk about individual goodness? This claim could equally be aimed at the network analysis, but it has a response. Much of the apparent vagueness about such talk stems from the fact that it is vague what the content of the better life lottery relation is. But if there is residual apparent vagueness, the network analysis has a standard response: this vagueness is to be accounted for by the nonobviousness of the analysis.

I end with an opinion. Granted Lewis's views about reference, under favourable conditions the *WG* analysis provides significant support for the risk neutrality thesis because of the simplicity argument and the pull towards natural properties. Under these conditions, the network analysis also provides significant support to the risk neutrality thesis because of the push provided by the pressure to vindicate the individual goodness platitudes. As conditions become more unfavourable to the *WG* analysis, the risk neutrality thesis remains correct, but the network analysis provides a greater share of the support. And if Lewis's views about reference are wrong, the risk neutrality thesis is still correct but the support provided by the network analysis is even more important. The *WL* analysis is less successful, but even if many of the issues about fixing the zero have been left open, the tools *WL* provides for thinking about this topic are very important. Finally, this has all been about meaning. The main arguments of *WG* and *WL* are about ethics. But when they are expressed in terms of individual goodness measures, the network analysis only makes the conclusions of these important arguments more secure.[17]

17. Support for this project was provided by the Arts and Humanities Research Council, and is gratefully acknowledged.