2 'Utility'

2.1 Usefulness

'Utility', in plain English, means *usefulness*. In Australia, a ute is a useful vehicle.

Jeremy Bentham specialized the meaning to a particular sort of usefulness. 'By utility', he said, 'is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered.'[1] The 'Principle of Utility' is the principle that actions are to be judged by their usefulness in this sense: their tendency to produce benefit, advantage, pleasure, good, or happiness. When John Stuart Mill speaks of the 'perfectly just conception of Utility or Happiness, considered as the directive rule of human conduct', he is using 'Utility' as a short name for this principle.[2] 'The Greatest Happiness Principle' was another name for it. People who subscribed to this principle came to be known as utilitarians.

Benthamism entered economics in 1873, with the publication of W. S. Jevons's *Theory of Political Economy*. Jevons quoted Bentham's definition of 'utility' and announced: 'This perfectly expresses the meaning of the term in Economy.'[3]

But after Jevons's time, the meaning of 'utility' in economics shifted. The word came to refer not to the tendency of an object to produce good, but to the good an object produces. By a person's 'utility', economists came to mean not the person's usefulness in promoting good around her, but her own good. 'Utility' came to mean *good*. This meaning has since been overlaid by yet another, which I shall be describing later. But it still persists as one of the current meanings of 'utility'.

I cannot give an authoritative history of the shift in meaning. One difficulty is that the interpretation of an author's intentions is often debatable. For Jevons, 'utility' definitely meant usefulness in Bentham's sense.

Alfred Marshall, too, thought of utilities as useful properties of objects. He said, for instance:

As [man's] production of material products is really nothing more than a rearrangement of matter which gives it new utilities; so his consumption of them is nothing more than a disarrangement of matter, which diminishes or destroys its utilities.[4]

But another remark of Marshall's illustrates the difficulty of interpretation:

The *total utility* of a thing to anyone (that is, the total pleasure or other benefit it yields him) increases with every increase of his stock of it, but not as fast as his stock increases.[5]

I believe Marshall still meant usefulness by 'utility' here. His parenthesis means, I think, that the *amount* of usefulness a thing has is equal to the *amount* of pleasure or other benefit it yields. But this remark could also be read – wrongly, I think – as *identifying* utility with pleasure or other benefit. Presumably it is ambiguities like this that allowed the shift of meaning to proceed unnoticed. And as early as 1881, F. Y. Edgeworth was occasionally using 'utility' unambiguously in the shifted sense. He referred to 'that quantity which alone the rational unionist is concerned to increase – the *labourer's utility*'.[6] An employer might be concerned to increase the labourer's usefulness, but not a unionist. Edgeworth meant the labourer's good. But this was not his normal terminology. Where later economists would have used 'utility', Edgeworth (like Jevons) normally used 'pleasure'. I do not think the shifted usage became common till much later.

Till recently it occurred exclusively in economics. I should be surprised to find an occurrence in philosophy from much before 1960. Henry Sidgwick's *Methods of Ethics*, the *locus classicus* of utilitarianism, hardly uses the word 'utility' at all. But it contains this footnote about its meaning:

I should point out that Hume uses 'utility' in a narrower sense than that which Bentham gave it, and one more in accordance with the usage of ordinary language. He distinguishes the 'useful' from the 'immediately agreeable': so that while recognising 'utility' as the main ground of our moral approbation of the more important virtues, he holds that there are other elements of personal merit which we approve because they are 'immediately agreeable', either to the person possessed of them or to others. It appears, however, more convenient to use the word in the wider sense in which it has been current since Bentham.[7]

Sidgwick says Bentham widened the sense of 'utility', whereas I said he narrowed it. I do not wish to quarrel about that. No doubt Sidgwick is right that to be immediately agreeable is not, in ordinary usage, a sort of usefulness. It is also true that some sorts of usefulness, according to ordinary usage, Bentham would not have included under 'utility' (the usefulness of a thumb-screw, for instance). So Bentham widened the meaning in one way and narrowed it in another. But the point is that agreeableness is the ten-

dency of an object to produce pleasure, not pleasure itself. So whether or not agreeableness is included in utility, utility is still a valuable tendency in an object, not a benefit derived from the object. Sidgwick, at this point in the book, is explaining that virtues have utility in that they 'are directly or indirectly productive of pleasure to ourselves or to others'.[8]

Recently, however, some philosophers have begun to adopt the economists' usage. This is unwise. As used by economists, the term 'utility' has become so ambiguous as to cause immense confusion. It should be used less, not more.

2.2   Axiomatic utility theory

The confusion stems from a new meaning that was assigned to the word as axiomatic utility theory developed during the course of the twentieth century.[9] The axiomatic theory sets out from a person's *preferences*. It proves that, provided these preferences conform to some axioms, they can be *represented* by a 'utility function'. The values taken by the function are called 'utilities'. The sense in which the function represents the preferences is this: of any pair of alternatives, the function assigns a greater utility to the one that is preferred. So 'utility' acquired the meaning: *the value of a function that represents a person's preferences*. This is by now the official definition of utility in economics. For brevity, let us say: utility is *that which represents a person's preferences*.

Now, let us ask this: of a pair of alternatives, is the one that a person prefers necessarily the one that is better for her? I mean nothing mysterious by this question. I use the word 'good' and its cognates – bad, better, best, and so on – in exactly the sense they have in ordinary conversation. Mother is using this sense when she tells you it would be good for you to have a few days' rest; an economist when she says inflation is worse for retired people than the unemployed; the politician when she tells you you would be better off dead than red. There is your good and my good: some things are good for you, and some are good for me. Everybody knows what 'good' means, though not many of us can define it. Of course, we endlessly disagree about what things are good and what things are bad. We argue with Mother, the economist, and the politician. One question we might argue over is whether it is necessarily good for a person to have what she prefers. That is the question I am asking now.

Again, then: of a pair of alternatives, is the one that a person prefers necessarily the one that is better for her? According to the official definition of 'utility', it has the greater utility. But a person's utility, as officially defined, has no necessary connection with her good. So nothing in the definition suggests that the preferred alternative is necessarily better for her. However,

many economists adopt the official definition of 'utility', whilst at the same time *also* using the word to stand for a person's good. Because an alternative preferred by a person is defined to have a higher utility for her, they take it for granted that it must be better for her. They suppose, then, that a person always prefers what is better for her.

We may call a person who always prefers what is better for herself 'self-interested'. This is using 'self-interested' in a very strong sense.[10] It is saying not only that the person pays no attention to the interests of other people, but also that she always prefers exactly what it is in her own interest. It rules out not only altruism, but also *imprudence*; a person is imprudent if, though concerned only for herself, she sometimes fails to do exactly what is best for herself. Simply by muddling the different meanings of utility', many economists find themselves committed to the view that people are necessarily self-interested in this strong sense.

'The first principle of Economics', said Edgeworth, 'is that every agent is actuated only by self-interest.'[11] That may have been true in Edgeworth's day, but it was one of the achievements of modern utility theory to free economics from such a dubious first principle. The achievement was announced by Lionel Robbins: 'So far as we are concerned', he said, 'our economic subjects can be pure egoists, pure altruists, pure ascetics, pure sensualists or – what is much more likely – bundles of all these impulses.'[12] The first principle of economics is, I take it, utility theory. And modern, axiomatic utility theory makes no assumption that people are self-interested. All it assumes is that people's preferences conform to a number of axioms: roughly, they simply need to be consistent. They can conform to the axioms without being self-interested. Yet the muddle over 'utility' leads many economists to forget this important discovery.

It is certainly not very plausible that people's preferences are always self-interested in the strong sense I described. It is a common opinion that many people – parents for instance – have preferences that are partly altruistic: directed towards the good of others. And it is a common opinion that many people are imprudent: for instance, they prefer to take less exercise than is good for them. If either of these opinions is correct, people's preferences are not self-interested.

It may turn out that common opinion is incorrect. Altruism and imprudence may not exist, and everyone may always prefer what is best for themselves. Utility, defined to represent a person's preferences, may indeed turn out also to represent her good. All this is arguable. One argument, for instance, is that a person's good actually *consists in* the satisfaction of her preferences, so that, of two alternatives, the one she prefers cannot fail to be better for her. But at least the argument needs to be made. It is a substantive question whether or not preferences are necessarily self-interested.

If, though, you use 'utility' to stand for a representation of a person's preferences, and at the same time for the person's good, you cannot even express the question. You will say: by definition, what a person prefers has more utility for her, so how can it fail to have more utility for her? The ambiguity is intolerable.

2.3   Expected utility theory

The confusion multiplies when it comes to expected utility theory, the branch of utility theory that takes account of uncertainty. Modern, axiomatic expected utility theory imposes axioms on a person's preferences between uncertain prospects – more axioms than ordinary utility theory does. Granted these axioms, the theory demonstrates the existence of a utility function that has two properties. First, the function represents the preferences, just as before: of two prospects, the preferred one has the higher utility. And, secondly, the function has the 'expected utility form', which means that the utility assigned to an uncertain prospect is the expectation (in probability theory's sense) of the utilities assigned to the prospect's possible outcomes. Axiomatic expected utility theory, then, is like ordinary axiomatic utility theory in that it defines utility to represent preferences. 'Utility' still means that which represents preferences; the person still maximizes her utility. But since the utility of a prospect is also its expected utility – the expectation of the utility of its outcomes – we can also say the person maximizes her expected utility.

It happens that utility defined this way is unique up to increasing linear transformations. Utility with this degree of uniqueness is often said to be 'cardinal'.

There is good evidence that people in practice do not conform to expected utility theory; they often violate its axioms.[13] But it can be argued (though this too is controversial) that fully rational people will conform to the theory.[14] For the sake of argument, let us take that for granted, and from now on consider only fully rational people. So we can assume they are expected utility maximizers. For the sake of argument, too, I now want to set aside the issue of self-interest I mentioned in section 2.2. So let us consider only people whose preferences happen to be self-interested, in the strong sense I described.

Take, then, a rational, self-interested person. When there is no uncertainty to worry about, this person prefers, of two alternatives, the one that is better for her. But what about her preferences between uncertain prospects? Of two prospects, will she necessarily prefer the one that gives her a greater *expectation* of good? Will she necessarily maximize her expected good?

A plausible answer is no, for two reasons. First, for an expectation of good even to exist, good must be an arithmetical quantity.[15] And it is plausible that good is not such a precise notion as that. It makes clear sense to say that one prospect is better or worse than another, so goodness at least constitutes an ordering (though not necessarily a complete one). But it is plausible that there are no precise arithmetical quantities of good.

And, secondly, even if there *are* arithmetical quantities of good, it is plausible that a rational, self-interested person might not maximize the expectation of her good. Suppose a person had a choice between ninety-nine units of good for sure, on the one hand, and, on the other, a gamble at equal odds between no units and two hundred units. The gamble has a higher expectation of good for her. Yet it seems perfectly rational for her to play safe and take the ninety-nine units for sure. She would do this if she was risk-averse about her good. Maximizing the expectation of good implies risk-neutrality about good. And it seems perfectly rational to take some different attitude, such as risk-aversion, to risk about one's good.

This needs some more explanation. Our subject is self-interested. Therefore, of two prospects, we can take it that she will prefer the one that is better for her. But it does not follow that she will prefer the one with the greater *expectation* of good for her. The one that is better for her may actually have a lower expectation of good for her. The example shows this. Though the option of ninety-nine units for sure has a lower expectation of good for her, it may nevertheless be better for her, because it is safe.

To be sure, since we are assuming the person conforms to expected utility theory, she maximizes the expectation of her utility. This means she is risk-neutral about *utility*. But it does not follow that she is risk-neutral about good. Axiomatic expected utility theory does not imply risk-neutrality about good; it does not imply that a rational person maximizes the expectation of her good.

But once again the ambiguity of 'utility' can get in the way of understanding this point. A rational person necessarily maximizes the expectation of utility. For an economist who also uses 'utility' to mean good, this will make it seem as though a rational person necessarily maximizes the expectation of her good. But this is a mistaken deduction.

It may turn out, on further investigation, that actually a rational self-interested person *does* necessarily maximize the expectation of her good. This is a common view that began, I believe, with Daniel Bernoulli.[16] I call it 'Bernoulli's hypothesis'. There are arguments in its favour.[17] But it is not implied by axiomatic expected utility theory. And we certainly need to be able to ask whether or not it is true. If, however, you define utility, on the one hand, so that a person necessarily maximizes the expectation of it, and,

on the other, you covertly identify utility with a person's good, then you cannot ask the question.

Let me express the question differently. Our self-interested subject always prefers, of two alternative prospects, the one that is better for her. Since she prefers it, by definition it has a higher utility. So her utility represents her good in the sense that, of any two prospects, the one that is better for her has the higher utility. This means that utility can properly be called an *ordinal* representation of her good. It has this property simply because the person is self-interested. But if she *also* maximizes the expectation of her good, then expected utility theory tells us that utility will represent her good more tightly than this: it will be an increasing linear transform of her good. In that case, it is said to represent her good *cardinally*. So Bernoulli's hypothesis is equivalent to the claim that utility represents good cardinally. Our question can be put this way, then: for a rational self-interested person, does her utility, as defined by expected utility theory, represent her good cardinally? We know already that utility itself is cardinal; it is unique up to increasing linear transformations. But that does not imply it represents good cardinally. The question is: actually, does it?

Put this way, it is a question that has very much interested welfare economists. A cardinal representation of good is very useful thing to have if you are interested in evaluating distributions of income and wealth. It is essential if you are a utilitarian concerned to maximize the total of people's good. It was very much in demand in the early 1950s. At that time, welfare economists had been deprived by the 'ordinalist revolution' of the 1930s of their right both to cardinal representations of good and to interpersonal comparisons of good. This had left them with almost nothing to say about distributions of income and wealth. The advent of expected utility theory in the 1940s appeared to offer them back at least the cardinal representations. So they urgently needed to know whether the offer could be trusted. Could expected utility theory really supply cardinal representations of good?

There was great confusion about the question at the time, fuelled by the ambiguity of 'utility'. The confusion showed up in a translation published by *Econometrica* in 1954 of Bernoulli's seminal article on expected utility theory. Bernoulli said that a rational person would maximize the expectation of her *emolumentum*. (He wrote in Latin.) '*Emolumentum*' means benefit or advantage. So this is a statement of Bernoulli's hypothesis, as I called it: a rational person maximizes her expectation of good. But in *Econometrica,* '*emolumentum*' was translated as 'utility'. This lapse in scholarship prevented readers from seeing the crucial difference between Bernoulli's version of expected utility theory and the axiomatic version. The axiomatic version is not committed to the view that a rational person

maximizes the expectation of her good. Bernoulli, on the other hand, was.

A clearsighted article published by Daniel Ellsberg in 1954 ought to have sorted out the muddle. Ellsberg, like me, was concerned about the ambiguity of 'utility'. Of von Neumann and Morgenstern, he says:

The operations that define their concepts are essentially new, and their results are neither intended nor suited to fill the main functions of the older, more familiar brands of 'cardinal utility'. It is unfortunate that old terms have been retained, for their associations arouse both hopes and antagonisms that have no real roots in the new context.[18]

I am sorry to say, however, that confusion persists, still fuelled by the same ambiguity. It is at work, for instance, in this more recent passage from John Harsanyi:

To be sure, the vNM utility function[19] of any given individual is estimated from his choice behavior under risk and uncertainty. But this does not mean that his vNM utility function is *merely* an indication of his attitude towards risk taking. Rather, as its name shows, it is a utility function, and more specifically, it is what economists call a cardinal utility function. This means that the primary task of a vNM utility function is *not* to express a given individual's attitude toward risk taking; rather it is to indicate how much utility, i.e. how much subjective importance, he assigns to various goals.[20]

2.4   Sen's usage

We cannot now return 'utility' to its original meaning of usefulness. It is a technical term thoroughly embedded in economics. But at least, as a technical term, we should confine it to one meaning. Which should we choose?

A natural place to look for leadership is the work of Amartya Sen. Let us examine how Sen uses this word 'utility' in his recent work. I shall take as examples his books *The Standard of Living* and *On Ethics and Economics*.

Sen's notion of utility plays an important role in his arguments. A major thesis of *The Standard of Living*, for instance, is that a person's standard of living cannot be identified with her utility. But Sen evidently assumes that his readers understand what he means by 'utility', and does not explain it well. I do not find his meaning obvious, though. Two things are clear. Sen does not mean by 'utility' what axiomatic utility theory means by it.[21] Nor does he mean a person's good; there are things he considers good for a person – for instance, her functionings and capabilities – that he does not include in her utility.

My best understanding of Sen's meaning for 'utility' is: *that which utilitarians believe to constitute good*. His reason for adopting this meaning, I suppose, is this. 'Utility' as a technical term was invented by utilitarians.

So we ought to give it the meaning they give it. And utilitarians intend 'utility' to refer to what they believe to constitute good. So we ought to use it to refer to that too.

But this is a poor reason. First of all, it is false that, generally, utilitarians intend 'utility' to refer to what they believe to constitute good. I said that in section 2.1. Among the classical utilitarian philosophers – Bentham, Mill, Sidgwick – none of them used the term that way. They intended it to refer to the tendency to promote good. Among the classical utilitarian economists – Jevons, Marshall, Edgeworth, Pigou[22] – only Edgeworth occasionally slipped into this usage. Even contemporary utilitarian philosophers rarely use 'utility' that way. Generally, utilitarians refer to those things they believe to constitute good by their specific names: 'pleasure', 'happiness', 'satisfaction', 'wellbeing', 'welfare', and so on.

Secondly, even if we suppose that utilitarians *do* intend 'utility' to refer to what they believe to constitute good, and even if we want to give 'utility' the meaning they give it, we should not use it to refer to the same thing as they do, unless we are utilitarians. (And Sen is not one.) In so far as utilitarians use this word in this way, they *mean* by it simply good. Of course, they intend it to *refer* to what they believe to constitute good. But if we are to give the word the same meaning as they do, we must use it to refer to what *does* constitute good. Only if utilitarians are right will this be just what utilitarians believe to constitute good.

If, then, Sen means by 'utility' that which utilitarians believe to constitute good, he is endowing the word with a new meaning. I believe this meaning may be unique to Sen. It is likely to be a useful one only if it is clear what utilitarians believe to constitute good; only then will its reference be unambiguous. But actually utilitarians are divided over what constitutes good. Some think good consists in good feelings; some in happiness; others in the satisfaction of desires; and so on.

Because of this ambiguity, Sen is constantly forced in these books to mention different 'conceptions of utility' separately: the happiness conception, the desire-satisfaction conception, and so on. His arguments have to deal separately with each. In showing that the standard of living is not the same as utility, for instance, he has to show it for the various conceptions one by one.[23] And, since the conceptions differ radically, each demands a radically different argument. I see, therefore, little point in collecting them together under the one heading of 'utility'. It would be a harmless thing to do if 'utility' were not already damagingly ambiguous. As it is, we can do without yet another meaning for it, and especially one that is itself ambiguous.

So I think Sen's lead points in the wrong direction.[24]

2.5   The best usage

As a meaning for 'utility', we should choose either *that which represents preferences*, on the one hand, or *good* on the other. Which should it be?

Both are economists' meanings. Philosophers have not used either till recently, and neither is well established in philosophy. So the needs of economics should have the first say in deciding between them. In any case, they will.

In economics the official meaning is the first: *that which represents preferences*. This is the meaning given in the major doctrinal texts[25] and the best textbooks.[26] It is defined with great precision, as a technical term should be. Its use is universal in theoretical economics. There is no alternative term. Economics cannot do without it.

'Utility' in this sense need not be confined to a representation of a person's actual preferences. A function can also be called a utility function if it represents the preferences a person would have if she were rational and self-interested. But if she were rational and self-interested, she would prefer, of two alternatives, the one that is better for her. So a function that represents the preferences she would have if she were rational and self-interested also represents, ordinally, her good. An ordinal representation of good can therefore be called a utility function.[27] We might also have a use for a notion of 'social' preferences, suitably interpreted, and social preferences too can be represented by a utility function. All of this is within the scope of the official definition of 'utility'.

Ellsberg thinks it unfortunate that 'utility' acquired this new technical meaning.[28] I disagree. Once divorced from usefulness, the word has no *natural* meaning. So it is ideally suited to perform the services of a technical term.

The second meaning, *good*, is also the property of economists. But it is an underground one. You will not often find it openly acknowledged. There is no classical warrant for it. Because it hides underground, its main effect has been to cause confusion. And – a further sin – it is perfectly redundant. We already have an excellent word with the meaning of good: 'good'.

I therefore propose that this second meaning for 'utility' should be prohibited. 'Utility' should be used only for a representation of preferences.

3   Extended preferences

3.1   Introduction

Many economists have adopted a doctrine known as 'ordinalism'. Ordinalism insists that we can know about people's good only by means of our knowledge of people's preferences. Many ordinalists believe their doctrine implies that we cannot know how one person's good compares with another's. But some have resisted this conclusion; they have argued that we can make interpersonal comparisons of good in a way that is consistent with ordinalism. They think a particular class of preferences can constitute a basis for comparisons between the good of different people. They have in mind people's preferences between very widely defined alternatives, each of which consists of a way of life together with the personal characteristics of a person who lives that life. These are called 'extended preferences'. This chapter reveals a flaw in the argument that extended preferences can be a basis for interpersonal comparisons of good. It shows that interpersonal comparisons really are inconsistent with ordinalism.

Section 3.2 of this chapter explains ordinalism in more detail. Section 3.3 describes the notion of extended preferences, and specifies a condition extended preferences must satisfy if they are to serve the purpose they are meant for: everyone must have the same extended preferences. Section 3.4 quotes an argument of John Harsanyi's that is intended to show everyone will indeed have the same extended preferences. Section 3.5 explains that Harsanyi's argument contains an error.

There is an alternative approach to interpersonal comparisons implicit in some ordinalist writings, including Harsanyi's. This second approach is not usually clearly distinguished from the approach through extended preferences, but actually it does not depend on extended preferences.

Contents