Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism
Author(s): John Broome
Source: *The Review of Economic Studies*, Vol. 57, No. 3 (Jul., 1990), pp. 477-502
Published by: Oxford University Press
Stable URL: http://www.jstor.org/stable/2298025
Accessed: 17/07/2011 16:50

Review of Economic Studies (1990) **57**, 477-502
0034-6527/90/00290477$02.00

Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism

JOHN BROOME
University of Bristol

First version received July 1987, *final version accepted October* 1989 (*Eds*)

This paper introduces the Bolker-Jeffrey version of expected utility theory, which differs in several important respects from the versions commonly used by economists. Within the Bolker-Jeffrey theory, the paper proves a theorem first proved by Harsanyi: if social preferences are coherent and Paretian, and individual preferences are coherent, then social utility can be taken to be the sum of individual utilities. But the paper shows that in the Bolker-Jeffrey theory the proof requires very stringent assumptions. It assesses the significance of this fact.

1. INTRODUCTION

In 1955, John Harsanyi proved a remarkable theorem. Assume everybody has preferences that conform to expected utility theory. And assume there are social preferences that also conform to expected utility theory. Finally, assume that the social preferences satisfy the Pareto criterion. Harsanyi proved that, given these three assumptions, social preferences can be represented (in the manner of expected utility theory) by a utility function that is the sum of utility functions representing the preferences of the individuals. I call this the "Utilitarian Theorem". Section 2 of this paper describes it in more detail.

The significance of the Utilitarian Theorem has been much debated (e.g. Hammond (1987), Jeffrey (1971), Sen (1976). Harsanyi believes it supports utilitarianism (Harsanyi (1977)). That is, perhaps, an overstatement, but I do think it throws enough light on the foundations of utilitarianism to justify the name I give it. It certainly makes a remarkable link between attitudes to risk and attitudes to inequality, which was Harsanyi's original purpose (Harsanyi (1953, 1955)). But this paper is not about the theorem's significance; I have expressed my own views on that elsewhere (Broome (1987, 1990, 1991)). It is about its truth.

A number of proofs have been published besides Harsanyi's (e.g. Border (1985), Deschamps and Gevers (1979), Fishburn (1984), Hammond (1981, 1983)). Each is tied to a particular version of expected utility theory, and several rely implicitly on strong assumptions. Harsanyi's own proof assumes that probabilities are objective and known to everyone. Other existing proofs allow for subjective probabilities, but they all assume versions of expected utility theory that derive ultimately from Savage's (1972). This paper tries out the theorem in the Bolker-Jeffrey version, which is radically different from Savage's. I shall argue in Section 3 that there are good reasons to test the theorem in this version. This paper proves the theorem within it. But it also shows the need for stringent assumptions.

One of my main aims is to introduce the Bolker-Jeffrey theory itself. It has some important attractions compared with Savage's theory, and deserves to be better known amongst economists. Section 4 of this paper describes the theory. But no presentation of the Bolker-Jeffrey theory these days can ignore the strong objections that have recently been raised against it from the direction of "causal" decision theories (which include Savage's theory). Section 5 explains these objections, using some economic examples. It also argues that, although they may be cogent objections to the theory conceived as a theory of decision, conceived as a theory of valuation they leave it unscathed. The Bolker-Jeffrey theory remains particularly appropriate for Harsanyi's theorem, which is best understood as a matter of valuation rather than decision.

Section 6 outlines the proof of the Utilitarian Theorem within Bolker-Jeffrey theory. This proof requires strong assumptions. Section 7 discusses the assumptions, and draws conclusions.

Appendix A contains counterexamples to the Utilitarian Theorem, demonstrating the need for the assumptions I mentioned. Appendix B contains the theorem's full proof. I hope the appeal of the methods associated with the Bolker-Jeffrey theory, displayed in these appendices, may help to recommend it to some readers.

2. THE UTILITARIAN THEOREM

Let there by h people. Let each have preferences defined on a set of alternatives involving some uncertainty (the same set for each person). Person i's preferences I shall designate by the symbols $\succ_i$, $\succcurlyeq_i$ and $\approx_i$ in the usual way. Throughout this paper, I shall assume that each person's preferences are *coherent*, by which I mean that they satisfy the axioms of expected utility theory. I want to allow for different versions of the theory, with different axioms, so I cannot yet define coherence more exactly.

Expected utility theory shows that each person's preferences may be represented by a utility function defined on the domain of alternatives. These utility functions are *expectational*. By this I mean that if an alternative has uncertain results, its utility is the expectation of the utility of its possible results. Again, the precise meaning of "expectational" can only be defined within each version of the theory. A person's utility function is not unique; several expectational utility functions will represent her preferences equally well. But all of them will be positive linear transforms, or (in the Bolker-Jeffrey theory; see Section 4) fractional linear transforms, of each other.

Let there also be social preferences defined on the same set of alternatives. I shall designate them by $\succ_g$, $\succcurlyeq_g$ and $\approx_g$. I call them *Paretian* if and only if, for all alternatives A and B,

if $A \approx_i B$ for all i then $A \approx_g B$, and

if $A \succcurlyeq_i B$ for all i and $A \succ_i B$ for some i then $A \succ_g B$.

If the social preferences are coherent they may be represented by an expectational utility function. Once again, many functions, all positive linear or fractional linear transforms of each other, will serve to represent the preferences. Let $U_1, \ldots, U_h$ be utility functions representing the individual's preferences, and U_g a utility function representing social preferences. Then if social preferences are Paretian, U_g will be a function of the U_i's:

$$U_g(A) = W(U_1(A), \ldots U_h(A)) \quad \text{for all alternatives } A.$$

And W will be increasing in each argument.

I call the social preferences *utilitarian* if and only if there is an expectational utility function U_g representing social preferences and for each i an expectational utility function U_i representing i's preferences such that

$$U_g(A) = \Sigma_i U_i(A) \quad \text{for all alternatives } A.$$

Now I can state the theorem that is the subject to this paper:

The Utilitarian Theorem. *Suppose that each person has coherent preferences. Then if social preferences are coherent and Paretian, they are utilitarian.*

Harsanyi's orginal proof of the Utilitarian Theorem took probabilities to be objective and known to everyone. Other proofs, however (including mine in this paper) are more general in that they allow for subjective probabilities that may differ from person to person. However, these proofs invariably come to the Utilitarian Theorem via a proof of the following

Probability Agreement Theorem. *Suppose that each person has coherent preferences. Then if social preferences are coherent and Paretian, the individual and social preferences must all agree about the probabilities they assign to every event.*

So the initial extra generality cancels itself out: probabilities have to be universally agreed anyway.

Nevertheless, the generality achieves something. It is better than simply assuming agreement about probabilities from the start. The Probability Agreement Theorem is important in its own right. It shows that the coherence and Paretian requirements on social preferences are together very stringent. They impose conditions, not just on the social preferences themselves, but on individual preferences too. As a general rule, we have no reason to expect individual preferences to agree about probabilities. Unless they do, though, the theorem says that social preferences cannot be both coherent and Paretian. Yet it is natural to think they should be both. Furthermore, coherence and the Paretian requirement are the conditions of the Utilitarian Theorem. So the Probability Agreement Theorem tell us that as a general rule the conditions of the Utilitarian Theorem are mutally inconsistent. It tells us, then, that some work of interpretation needs to be done in order to reconcile the two conditions. Without this work, neither the notion of social preferences, nor the Utilitarian Theorem, can be properly understood. I have attempted it myself in Broome (1987) and (1989).

In summary, the theorems say that unless individuals agree about probabilities there can be no coherent Paretian social preferences. And when coherent Paretian social preferences do exist, they must also agree about probabilities, and they must be utilitarian.

3. THE EX POST APPROACH

Existing proofs of the Utilitarian Theorem—those that allow for subjective probability (Deschamps and Gevers (1979), Hammond (1981, 1983))—model uncertainty in a way that is, broadly speaking, Leonard Savage's (1972). In this model there are a number of "states of nature", any one of which may come about. People have preferences between alternative prospects ("acts" in Savage's terminology). Each prospect associates a particular outcome ("consequence" in Savage's terminology) with each state of nature: if the prospect is chosen and the state comes about then this outcome will result. This structure is shown in Table 1. The cells of the table show outcomes A_1, A_2, B_1 and so on.

TABLE 1

	States of nature			
Prospects	S_1	S_2	S_3	...
A	A_1	A_2	A_3	...
B	B_1	B_2	B_3	...
C	C_1	C_2	C_3	...
⋮	⋮	⋮	⋮	

Each person has preferences amongst the alternative prospects. Expected utility theory tells us that, provided they are coherent, these preferences may be represented by probabilities and utilities. Probabilities are attached to the states of nature; utilities, initially, to outcomes. Derivatively each prospect has a utility too, calculated as the expectation of the utility of its possible outcomes, assessed according to the probabilities. So the utility of prospect A is

$$U(A) = \mu(S_1)U(A_1) + \mu(S_2)U(A_2) + \cdots$$

where μ stands for probability. Of two prospects, the preferred one will have the higher utility.

The utility of a prospect for a person, then, is derived from the utilities of outcomes, and it depends on the person's assessment of probabilities. On the other hand, in the model the utility of an outcome is basic. This leaves open an escape route from the Utilitarian and Probability Agreement Theorems. Many people find the implications of these theorems unattractive. It seems desirable to have coherent social preferences. But according to the theorems this is rarely possible, and even when it is, the social preferences must be utilitarian in the sense defined in Section 2. Such utilitarian preferences seem, at least at first sight, to deny the value of equality. Whatever the truth about this—and in this paper I am not going to enquire into it (see Broome (1990))—an escape from these implications is to adopt the so-called "ex post" approach to forming social preferences. This approach has been proposed by Hammond (1983) among others, and I have argued for it myself (Broome (1982)).

The idea of the ex post approach is that social preferences about prospects should be based on individuals' preferences about the possible outcomes of those prospects, but not necessarily on their preferences about the prospects themselves. So one should require social preferences to be Paretian about outcomes but not about prospects generally. If everyone's preferences assign one outcome a higher utility than another, then so should social preferences, but the same need not be true for prospects. With this looser Paretian requirement, coherent social preferences are easier to come by, and they need not be utilitarian.

The argument for the ex post approach is that people's preferences about prospects do not depend only on their wants, but also on their beliefs about probabilities. Democratic principles may insist that social preferences should be based on people's wants, but it is quite a different matter to insist that they should be based on their beliefs too.

But this argument is open to a powerful objection. People's preferences about prospects doubtless depend on their beliefs as well as their wants. But so do their preferences about anything. The ex post approach assumes that outcomes can be distinguished from prospects in such a way that preferences about outcomes do not depend on beliefs about probabilities. But it is never certain what good or harm can result from

anything. So a person's preferences about anything must depend on her beliefs about the probabilities of its possible results. Take, for instance, one of Savage's (1972, p. 25) examples of a "consequence" (outcome): a refreshing swim with friends. If I have a refreshing swim with friends I might or might not get cramp, and my preferences about the swim will depend on my beliefs about the probabilities of these results. If I swim and get cramp, I might or might not drown, and my preferences about swimming and getting cramp will depend on my beliefs about the likelihood of these results. And so on. No doubt in a practical decision-making problem of the sort Savage was concerned with, it is often possible to draw a workable distinction between prospects whose value depends on the probabilities of their results, and outcomes that have value in their own right. But this distinction, the objection goes, cannot be sustained in principle. And the ex post approach cannot be justified without it.

I think this is a strong objection but not necessarily a conclusive one. It may be that actually an appropriate distinction can be drawn between prospects and outcomes. For instance, complete possible worlds are plausible candidates for outcomes. But the objection certainly needs to be taken seriously. We should therefore not rely on conclusions drawn from a version of expected utility theory that takes for granted the distinction between prospects and outcomes.

The Bolker-Jeffrey theory—the idea and interpretation is Richard Jeffrey's (1983), the axiomatization Ethan Bolker's (1966, 1967)—assumes no such distinction. Indeed Bolker's axiomatization explicitly rules it out, as I shall explain. A good reason, therefore, for trying out the Utilitarian Theorem within this theory is that, in so far as the theorem is true within it, no escape to an ex post approach is available.

4. INTRODUCTION TO THE BOLKER-JEFFREY THEORY

In Bolker-Jeffrey expected utility theory, preferences, utilities and probabilities are all defined on the same set of prospects. Jeffrey expresses these prospects as *propositions*, such as "I have a refreshing swim with friends", and applies the propositional calculus to them. A prospect or proposition may be thought of as a subset of the set of all possible worlds, the subset consisting of worlds where the proposition is true. The operations of propositional calculus correspond to set-theoretic operations. If A is a proposition, $\sim A$ (i.e. not A) is the complement of A. $A \vee B$ (i.e. A or B) is the union of A and B; $A \wedge B$ (i.e. A and B) their intersection. If A and B are contraries (propositions that cannot both be true) they are disjoint sets.

Let A be "I have a refreshing swim with friends", B "I get cramp", and C "I drown". Let A_1 be $A \wedge B$ and let A_2 be $A \wedge \sim B$. Then $A = A_1 \vee A_2$ (the disjunction of A_1 and A_2), and A_1 and A_2 are contrary propositions. Rules I shall describe later say that, whenever A is the disjunction of two contraries A_1 and A_2,

$$\mu(A) = \mu(A_1) + \mu(A_2)$$

and

$$U(A) = \frac{\mu(A_1)U(A_1) + \mu(A_2)U(A_2)}{\mu(A_1) + \mu(A_2)} = \frac{\mu(A_1)}{\mu(A)}U(A_1) + \frac{\mu(A_2)}{\mu(A)}U(A_2).$$

The formula for probability is obviously appropriate. To understand the formula for utility, remember that $\mu_1(A)/\mu(A)$ and $\mu_2(A)/\mu(A)$ are the probabilities of A_1 and A_2 conditional on A. So the formula says that the utility of A is the expectation of utility given that A is true. In this way utility is expectational.

A feature of Bolker's axiomatization is that the set of prospects is *atomless*. This means that any prospect in the set can always be broken down into a disjunction in the way that A breaks down into A_1 and A_2. A_1, for instance, breaks down into $A_{11} = A_1 \wedge C$ and $A_{12} = A_1 \wedge \sim C$, so that $A_1 = A_{11} \vee A_{12}$. And

$$U(A_1) = \frac{\mu(A_{11})}{\mu(A_1)} U(A_{11}) + \frac{\mu(A_{12})}{\mu(A_1)} U(A_{12}).$$

The assumption is that any prospect breaks down similarly into a disjunction, and has its utility resolved into the expectation of the utilities of its disjuncts. The utility of any proposition, then, depends on probabilities and utilities, on beliefs as well as wants. There are no propositions that play the role of outcomes or consequences and have a utility that is independent of probability judgments. The ex post approach is therefore not possible within the Bolker–Jeffrey theory.

A major difference between the Bolker–Jeffrey theory and others is this. In the Bolker–Jeffrey theory, when prospects are combined together by the truth-functional operation of disjunction they retain their own probabilities, made conditional on the disjunction. Other theories combine prospects or outcomes by forming gambles. This involves artificially assigning a probability (or a state of nature, which has its own probability) to each outcome. Setting up a gamble, in fact, involves altering causal relations in the world. In practice the sorts of gamble that are required by the theory may be causally impossible. For instance, fine weather tomorrow may be assigned to the state of nature: this coin falls heads on its next toss. But the toss of a coin cannot actually determine what the weather will be. In order to include them in her preference ordering a person has to imagine herself being offered such impossible gambles. The Bolker–Jeffrey theory, on the other hand, assumes that a person retains her actual beliefs about the causal processes in the world. Jeffrey (1983, p. 157) considers this the theory's main advantage.

Take a set $\mathcal{S}$ of prospects that is closed under the operations of disjunction and negation (or union and complementation):

If A and B are in $\mathcal{S}$ then $\sim A$ and $A \vee B$ are in $\mathcal{S}$.

$\mathcal{S}$ is then a *Boolean algebra*. (For an account of Boolean algebras see Sikorski (1960).) It will contain a unit T and a zero F. T is the necessarily true proposition and F the necessarily false one:

$$T = A \vee \sim A \quad \text{for all } A \text{ in } \mathcal{S}$$

$$F = \sim T.$$

T is the set of all possible worlds and F the empty set.

The Bolker–Jeffrey theory takes the field of preferences to be a Boolean algebra $\mathcal{S}$ with the zero removed. Write this field $\mathcal{S}' = \mathcal{S} - \{F\}$. F, then, has no place in the preference ordering.

The algebra $\mathcal{S}$ need not contain every set of possible worlds. In fact we assume that $\mathcal{S}$ is *atomless*, which rules this out. An atomless Boolean algebra is one whose every element has a non-zero strict sub-element:

For each $A \in \mathcal{S}$ other than F there is a $B \in \mathcal{S}$ such that $B \to A$ and $B \neq A$ and $B \neq F$.

This implies, for one thing, that the algebra cannot contain a set consisting of a single world. I have already described the significance of this assumption of atomlessness. We

also assume that the algebra is *complete*. This means that it contains all disjunctions of arbitrary sets of contrary members (see Sikorski (1960 p. 58)).

Like other expected utility theories, the Bolker-Jeffrey theory starts from given preferences and shows that, provided these preferences satisfy certain axioms, they can be represented by an expectational utility function. The axioms are: that the preferences are a complete preorder on $\mathscr{S}'$, that they are continuous in a particular sense, and that they satisfy these two conditions:

(i) (Averaging) if A,B in $\mathscr{S}'$ are contraries then

$$A \succ B \text{ implies } A \succ (A \vee B) \succ B$$

$$\text{and } A \approx B \text{ implies } A \approx (A \vee B) \approx B$$

(ii) (Impartiality) if A,B and C in $\mathscr{S}'$ are pairwise contraries, and

$A \approx B$ but not $A \approx C$, and $(A \vee C) \approx (B \vee C)$, then for every

D in $\mathscr{S}'$ that is contrary to A and B, $(A \vee D) \approx (B \vee D)$.

The averaging axiom says that a disjunction lies somewhere between the disjuncts in the preference ordering. It slightly resembles the independence axiom found in other versions of expected utility theory. The independence axiom implies that a "probability mixture" of two prospects lies somewhere between the prospects in the preference ordering. But it implies much more than this too, and furthermore it implies it for *any* probability mixture with any arbitrary probabilities. As I explained earlier, in Bolker-Jeffrey theory, on the other hand, prospects always carry their own probabilities with them when they combine in a disjunction. Combination with arbitrary probabilities is not allowed. So the averaging axiom is much weaker than independence. The much-criticized independence axiom is not required by Bolker-Jeffrey theory.

The impartiality axiom is less transparent. Take two contrary propositions A and B that are indifferent to each other. Form their disjunctions with a third contrary proposition that is not indifferent to them. The disjunctions $A \vee C$ and $B \vee C$ will be indifferent to each other if and only if A and B are equally probable. So a way of testing whether two indifferent propositions are equally probable is to compare together the disjunctions they form with a third, non-indifferent, proposition. The impartiality axiom says this test will deliver the same answer whatever third non-indifferent propositions is used.

Compare Savage's Postulate 4 (Savage (1972, p. 31)). Savage, too, needs to test whether two events, say E and F, are equally probable. He does this by taking a pair of outcomes, say A and B, that are known not to be indifferent. He forms a gamble $(A, E; B, F)$ in which A comes about in event E and B in F. And he forms the opposite gamble $(B, E; A, F)$. The events are equally probable if and only if these gambles are indifferent. Savage's Postulate 4 says that this test will deliver the same answer whatever pair of non-indifferent outcomes A and B are used.

The impartiality axiom is unsatisfactory in one respect. The explanation I have given for it presupposes expected utility theory to some extent. I said that $A \vee C$ and $B \vee C$ will be indifferent (for indifferent A and B and non-indifferent C) if and only if A and B are equally probable. But the reason for this is that the utility of a disjunction is the average of the utility of the disjuncts, weighted by their probabilities. And this reason comes out of expected utility theory. It is unsatisfactory that an axiom from which expected utility theory is supposed to be derived needs to be explained in this way.

Savage's Postulate 4 is in exactly the same position. The gambles $(A, E; B, F)$ and $(B, E; A, F)$ will be indifferent if and only if E and F are equally probable. The reason for this is that the utility of a gamble like this is the average utility of the outcomes, weighted by the probabilities of the events. This reason comes out of expected utility theory. But the postulate is one of the axioms from which expected utility theory is supposed to be derived. The impartiality axiom, then, is neither more nor less unsatisfactory than Savage's.

Now we come to the representation theorem:

Bolker's Existence Theorem. *Let $\mathcal{S}$ be a complete atomless Boolean algebra, and let $\succcurlyeq$ be coherent preference on $\mathcal{S}'$. Then there is a probability measure μ on $\mathcal{S}$ and a signed measure ν on $\mathcal{S}$ such that for all A and B in $\mathcal{S}'$*

$$A \succcurlyeq B \quad \text{if and only if } \frac{\nu(A)}{\mu(A)} \geqq \frac{\nu(B)}{\mu(B)}.$$

(I shall use the unqualified term "measure" to include signed measures, non-negative measures and non-positive measures. By a "non-negative measure" I mean a measure μ such that $\mu(A) \geqq 0$ for all A in $\mathcal{S}$. By a "positive measure", I mean a non-negative measure μ such that $\mu(A) = 0$ implies $A = F$.)

Because μ and ν are measures, whenever A and B are contraries (disjoint) then

$$\nu(A \vee B) = \nu(A) + \nu(B)$$

and

$$\mu(A \vee B) = \mu(A) + \mu(B).$$

In the existence theorem the role of utility is played by the quotient of measures ν/μ. Granted the existence of μ and ν, we define utility U on $\mathcal{S}'$ by

$$U(A) = \frac{\nu(A)}{\mu(A)} \quad \text{for all } A \text{ in } \mathcal{S}'.$$

U will then be a properly expectational utility function as required. For if A and B are contraries

$$U(A \vee B) = \frac{\nu(A \vee B)}{\mu(A \vee B)} = \frac{\nu(A) + \nu(B)}{\mu(A) + \mu(B)} = \frac{\mu(A)U(A) + \mu(B)U(B)}{\mu(A) + \mu(B)}. \tag{3}$$

I explained, using the example above, that this formula makes $U(A \vee B)$ the expectation of utility given that $A \vee B$ is true. The measure ν is best thought of as a convenient construction, the product of utility and probability.

Notice that the averaging condition rules out propositions (other than F) that have probability zero. For if A, $B \in \mathcal{S}'$, $A \succ B$ and $\mu(B) = 0$, we shall have from equation (3) that $U(A \vee B) = U(A)$, contrary to the averaging condition. The measure μ is therefore strictly positive (so U is well-defined on $\mathcal{S}'$). Bolker (1967, p. 337) defends this assumption by saying that propositions to which a person assigns probability zero can simply be left out of $\mathcal{S}$. This may be all right for a single person. But in Section 6, I shall be assuming that everyone's preferences and social preferences are defined on the same field $\mathcal{S}'$. So there is a substantive assumption implied here: that all these different preferences attach zero probability to the same set of propositions.

Notice too that, for a similar reason, the averaging condition rules out infinite values for $\nu(A)$.

Bolker's Uniqueness Theorem. *Let μ, μ' be probability measures and ν, ν' signed measures on a complete atomless Boolean algebra $\mathcal{S}$. Then μ, ν represent the same preferences as μ', ν' if and only if*

$$\left.\begin{aligned} \nu' &= a\nu + b\mu \\ \text{and} \\ \mu' &= c\nu + d\mu \end{aligned}\right\} \tag{4}$$

where

$$\left.\begin{aligned} ad - bc &> 0 \\ c\nu(T) + d &= 1 \\ \text{and} \\ c\nu(A) + d\mu(A) &> 0 \quad \textit{for all } A \textit{ in } \mathcal{S}'. \end{aligned}\right\} \tag{5}$$

The transformation of μ, ν to μ', ν' transforms utility U to

$$U' = \frac{\nu'}{\mu'} = \frac{a\nu + b\mu}{c\nu + d\mu} = \frac{aU + b}{cU + d}. \tag{6}$$

This is a fractional linear transformation. The Bolker-Jeffrey theory, then, allows a wider range of transformations for utility than other expected utility theories. It also allows transformations of probabilities, which other theories rule out.

Bolker (1967, 1974) and Jeffrey (1983, p. 161) have given an explanation of why other theories determine utilities and probabilities more tightly than theirs does. Other theories use a richer body of preferences as data on which to construct utilities. Preferences are defined on all prospects in which outcomes are assigned to arbitrary states of nature or assigned arbitrary probabilities. As I explained, such gambles may be causally impossible. In the Bolker-Jeffrey theory prospects always retain their own probabilities; when prospects are combined in a disjunction the probabilities are simply made conditional. It turns out that utilities and probabilities are then more loosely determined.

There is one case, however, where probabilities cannot be transformed and utilities are confined to a positive linear transformation only. This is where the range of U on $\mathcal{S}'$ is unbounded above and below. Condition (5) requires that $cU(A) + d > 0$ for all A in $\mathcal{S}'$. If U is unbounded above and below this is possible only if $c = 0$. Then by (5) $d = 1$, and by (6) $U' = aU + b$. So any transform of U is also unbounded above and below. I shall call preferences *unbounded* if they are represented by a utility function that is unbounded above and below.

Other decision theories rule out unbounded utilities as impossible. In effect, this is because of the St. Petersburg Paradox. Given a sequence of prospects with unbounded utilities, one can construct out of them a gamble that has infinite utility. This is what the St. Petersburg game does. And an infinite utility cannot be accommodated within the theory. But to construct such a gamble one needs to assign each outcome an artificially chosen probability. And this, as I have explained, is not allowed in the Bolker-Jeffrey theory. So unbounded utilities are not ruled out (see Jeffrey (1983, pp. 150-5)).

Indeed, in the Bolker-Jeffrey theory utility functions that are unbounded *either* above or below are nothing out of the ordinary. They can always be transformed into bounded functions by a suitable choice of coefficients in equation (4). And any utility function that does not attain its upper or lower bound can be transformed into one that is unbounded

above or below (Jeffrey (1983, p. 106)). This makes it clear that unboundedness need not imply extreme desirability, whatever that might mean.

Functions that are unbounded above *and* below are in a different class, because they cannot be transformed into bounded functions and they permit no transformation of probability. But one thing that is missing, I think, from Bolker's and Jeffrey's accounts of their theory is a characterization of unbounded preferences. What must be special about preferences to make their utility representation unbounded above and below? One feature they must have is to possess no top and no bottom: no prospect preferred or indifferent to every prospect and no prospect to which every prospect is preferred or indifferent. But it is not only unbounded preferences that have this feature. And it is not clear what extra feature unbounded preferences must have. Jeffrey (1983, p. 142) gives a necessary and sufficient condition for preferences to be unbounded, but it is intuitively opaque.

Throughout this paper I shall adopt the convenient normalization that

$$\nu(T) = \nu'(T) = 0.$$

Since μ is a probability measure

$$\mu(T) = \mu'(T) = 1.$$

From (4)

$$\nu'(T) = a\nu(T) + b\mu(T)$$

and

$$\mu'(T) = c\nu(T) + d\mu(T).$$

So $b = 0$ and $d = 1$. Then (4), (6) and (5) become

$$\left.\begin{aligned} &\nu' = a\nu \\ &\mu' = c\nu + \mu \\ &\text{and} \\ &U' = \frac{aU}{cU+1} \end{aligned}\right\} \tag{7}$$

where

$$\left.\begin{aligned} &a > 0 \\ &\text{and} \\ &c\nu(A) + \mu(A) > 0 \quad \text{for all } A \text{ in } \mathscr{S}'. \end{aligned}\right\} \tag{8}$$

I shall call a transformation *legal* if it meets these conditions.

5. DECISION VERSUS VALUATION

Consider this "twin prisoners' dilemma" (Lewis (1979)). You and your twin are facing a prisoners' dilemma. Table 2 shows the benefits (money, say): first in each bracket is your benefit, then hers.

TABLE 2

	Your twin	
You	Acts nice	Acts nasty
Act nice	(5, 5)	(0, 7)
Act nasty	(7, 0)	(1, 1)

There is no love lost between the pair of you, and you are going to act entirely self-interestedly. But she and you think very much alike, and you know this. So you know that if you act nice she will probably act nice too, and if you act nasty she will probably act nasty too.

Let A be the proposition "You act nasty" and B "Your twin acts nasty". You know that the probability of B given A, $\mu(A \wedge B)/\mu(A)$, is high. In Bolker-Jeffrey theory,

$$U(A) = U(A \wedge B)\mu(A \wedge B)/\mu(A) + U(A \wedge \sim B)\mu(A \wedge \sim B)/\mu(A).$$

So if $\mu(A \wedge B)/\mu(A)$ is high, $U(A)$ is near $U(A \wedge B)$. Similarly, if C is "You act nice" and D is "Your twin acts nice", $U(C)$ is near $U(C \wedge D)$. But $U(C \wedge D)$ is well above $U(A \wedge B)$ because $C \wedge D$ gets you 5 and $A \wedge B$ only 1. So $U(C)$ will be above $U(A)$.

This suggests you ought to act nice. But to most people (an exception is Horgan (1981)) this seems to be an incorrect conclusion. Whatever your twin does, you do better by acting nasty. Acting nasty is a dominant strategy, so that acting nice is irrational. Bolker-Jeffrey theory seems to come to the wrong conclusion in this case. Its mistake is fairly plain. Acting nice gives you evidence that your twin will probably act nice too, because she is like you. In a sense, then, it makes it probable that she will act nice, and so give you a good result. That is why the theory gives acting nice a high utility. But your acting nice does not have any causal influence on how she acts. And that is what counts in deciding what to do. The theory seems to have muddled evidence with cause.

Examples like this have led to a resurgence of "causal decision theory" (e.g. Gibbard and Harper (1978), Lewis (1981), Skyrms (1982)) in opposition to the "evidential" Bolker-Jeffrey theory. Savage's theory is the leading example of a causal decision theory. In Savage's theory there must be states of nature whose probabilities are independent of actions. Faced with the twin prisoners' dilemma, a follower of Savage has two alternatives. She may decline to apply the theory at all, perhaps taking the general view that decision theory does not apply to games. Or she may pick some things to serve as states of nature. She may, for instance, take your twin's acts as states of nature from your point of view. They will then have to be assigned probabilities independent of your own acts. And whatever probabilities they are assigned, acting nasty will come out with a higher expected utility for you. Either way, Bolker-Jeffrey theory's incorrect conclusion is avoided. So, on the face of it at least, we have here a weakness in Bolker-Jeffrey theory, and one that causal decision theory does not share.

It is a serious weakness, too, because situations like the twin prisoners' dilemma are common in practice. Many free-rider problems have the structure of a prisoners' dilemma. And in many of them the participants, though not twins, are similar enough for their behaviour to be quite closely correlated. For instance, suppose I, considering my own interest only, am wondering whether to join a union. If a lot of people join, we shall all be better off because we shall win benefits. But, however many people join, I should always be better off not joining, because I should get the benefits anyway, and save the

dues. I am much like other people, though. So if I join, many other people will probably do so too. The argument I gave for the twin prisoners' dilemma applies here too. Bolker-Jeffrey theory therefore seems to say I should join. But actually, most people would think, if I am concerned with my own interest only, I should not.

A government, too, may face dilemmas that raise the same difficulty. Take a government faced with a public that has rational expectations. It is wondering whether or not to expand the money supply. Table 3 shows how the effects of its actions will depend on what the people expect.

If the government expands the money supply, the people will probably have predicted that, so the result will be inflation. If it does not expand it, they will probably have predicted that too, so the result will be no change. Bolker-Jeffrey theory, then, will assign a higher expected utility to not expanding. It suggests this is the right thing to do. Dominance reasoning, however, shows that the right thing is to expand. That, at any rate, is the conclusion of most authors who have considered this "time-inconsistency problem" (e.g. Barro and Gordon (1983)). This government's dilemma has exactly the form of the "Newcomb problem" (Nozick (1969)), which first led to the renewed interest in causal decision theory. (The connection between the time-inconsistency problem and the Newcomb problem was noticed by Frydman, O'Driscoll and Schotter (1982).)

There may be ways for Bolker-Jeffrey theory to overcome the weakness I have been describing (see Eells (1982), Jeffrey (1987)). But in any case, it is only a weakness when the theory is taken literally as a theory of *decision*. If it is taken as a theory of *valuation*, there is no problem. Although, in the twin prisoners' dilemma, you ought to act nasty, it would nevertheless be *better* for you, in a natural sense, if you acted nice. If, for instance, you were to learn in some way that you were going to act nice, you would be justifiably pleased (see Jeffrey (1983, p. 82-83)). That news would tell you that your twin would probably do the same, so that the outcome would probably be a good one. In the other example, it would be better, in a natural sense, if the government kept the money supply constant. This is what the Bolker-Jeffrey theory says, and it makes good sense. It is, furthermore, a sense that causal decision theory cannot recognize. So here, as a theory of valuation, Bolker-Jeffrey theory has the advantage.

Valuation must be connected with decision in the end. The point of making valuations is to supply reasons for deciding one way or another when there is a decision to be made. But the connection need not be immediate. For instance, before it comes to a decision in some matter, a government may need to make complicated calculations about how good or bad the alternatives are for the people. In these calculations it may be appropriate for it to take "good" and "bad" for the people in a sense that is not directly connected with decisions the people make. So there is nothing wrong in general with separating value from decision. And it is useful in our context.

Let us ask: which sort of expected utility theory is it appropriate to apply to the Utilitarian Theorem—a theory of decision or a theory of valuation? The right one to

TABLE 3

	Expected action	
Government's action	Do not expand	Expand
Do not expand	No change	Depression
Expand	Increased employment	Inflation

pick is the one that makes the best sense of the theorem's assumptions: coherence and the Paretian assumption.

The Utilitarian Theorem is about individual and social *preferences.* The notion of preference is elastic, and we can interpret it as best fits the theorem. In one sense, a rational person in the twin prisoners' dilemma will prefer the prospect of acting nasty, since this is what she ought to do. In another sense, she will prefer the prospect of acting nice, since that would be better for her. Let us call these the "decisional" and "valuational" senses of "prefer". The same distinction applies to social preferences too. Of two alternatives, the one that is socially preferred in the decisional sense is the one the government should choose if it has the choice. The one that is socially preferred in the valuational sense is the better one. (This distinction is more thoroughly examined in Broome (1989).)

Suppose we pick the valuational interpretation. Bolker-Jeffrey theory suits this interpretation, and we can expect individual and social preferences, under this interpretation, to be coherent according to Bolker-Jeffrey theory. What about the Paretian assumption? Look again at the twin prisoners' dilemma. Under the valuational interpretation, you prefer the prospect of your acting nice to the prospect of your acting nasty. Obviously your twin prefers that too. So the Paretian assumption requires that your acting nice should be socially preferred to your acting nasty. And this seems quite right in the valuational sense: it would indeed be better if you acted nice. All the assumptions of the Utilitarian Theorem, then, work well under the valuational interpretation. This means that Bolker-Jeffrey theory suits the theorem.

Suppose, alternatively, we pick the decisional interpretation. Savage's theory suits this interpretation, and we can expect individual and social preferences, under this interpretation, to be coherent according to Savage's theory. What about the Paretian assumption? Well, I can find no way of setting up an example like the twin prisoners' dilemma so as to give worthwhile scope to the Paretian assumption. I said that Savage's theory can get a grip in the example by treating your twin's acts as states of nature from your point of view. Given that, your acts can be thought of as prospects, and acting nasty has a greater expected utility than acting nice. But to apply the Paretian assumption, we need both players to have preferences over the same prospects. This means we need to define states of nature that are the same for both. And however this is done, prospects will have to be joint acts: yours and your twin's. Your own acts will not be prospects on their own, so it will no longer be possible to say that your acting nasty has a greater expected utility for you. I conclude, then, that if we want to apply the Utilitarian Theorem to examples like this, the decisional interpretation will get us nowhere.

But the valuational interpretation works. I think this shows, at the very least, that this interpretation is as good as the decisional one for the purposes of the Utilitarian Theorem. But the objections I raised to the Bolker-Jeffrey theory earlier in this section only apply when it is taken as a theory of decision. Taken as a theory of valuation, Bolker-Jeffrey theory is in good shape. It is therefore appropriate to apply it to the Utilitarian Theorem, and Section 3 of this paper offered a good reason for doing so.

6. THE UTILITARIAN THEOREM IN THE BOLKER-JEFFREY THEORY

Let $\mathscr{S}$ be a complete atomless Boolean algebra. Let its unit be T and its zero F. Let $\mathscr{S}' = \mathscr{S} - \{F\}$. Let there be $h+1$ preference relations $\succcurlyeq_1, \succcurlyeq_2, \ldots, \succcurlyeq_h, \succcurlyeq_g$ on $\mathscr{S}'$. Write $I = \{1, \ldots, h\}$ and $I^+ = \{1, \ldots, h\} \cup \{g\}$.

Assumption 1. For all $i \in I$, $\succcurlyeq_i$ is coherent.

Assumption 2. $\succcurlyeq_g$ is coherent.

Given these assumptions, Bolker's Existence Theorem tells us that for each $i \in I^+$ there is a probability measure μ_i and a signed measure ν_i such that, for all A, B in $\mathcal{S}'$

$$\frac{\nu_i(A)}{\mu_i(A)} \geqq \frac{\nu_i(B)}{\mu_i(B)} \quad \text{if and only } \textit{if } A \succcurlyeq_i B.$$

I explained in Section 3 that μ_i is strictly positive and ν_i finite. For each $i \in I^+$ normalize ν_i to make $\nu_i(T) = 0$. Write $U_i(A) = \nu_i(A)/\mu_i(A)$.

Some definitions:

$\succcurlyeq_g$ is *Paretian* if an only if, for all A and B in $\mathcal{S}'$,

$$\text{if } A \approx_i B \text{ for all } i \in I \text{ then } A \approx_g B, \text{ and} \tag{9}$$

$$\text{if } A \succcurlyeq_i B \text{ for all } i \in I \text{ and } A \succ_i B \text{ for some } i \in I \text{ then } A \succ_g B. \tag{10}$$

$\succcurlyeq_g$ is *utilitarian* if and only if for each $i \in I^+$ there is a legal transform U_i' of U_i such that

$$U_g'(A) = \textstyle\sum_{i \in I} U_i'(A) \quad \text{for all } A \in \mathcal{S}'.$$

$\succcurlyeq_i$ is *unbounded* if and only if, for all $n = 1, 2, \ldots$ there are A^n and B^n in $\mathcal{S}'$ such that $U_i(A^n) > n$ and $U_i(B^n) < -n$.

I shall assume:

Assumption 3. $\succcurlyeq_g$ is Paretian.

Assumption 4. For all $i \in I$ there is an $A_i \in \mathcal{S}'$ such that $A_i \approx_j T$ for all $j \in I - \{i\}$ but not $A_i \approx_i T$.

Assumption 4 requires, firstly, that no one is entirely indifferent between all propositions. And secondly it requires some minimal independence between different people's preferences.

Theorem 1. *Under Assumptions* 1-4, *for each* $i \in I$ *there is a number* $e_i > 0$ *such that for all* A *in* $\mathcal{S}'$

$$\nu_g(A) = \textstyle\sum_{i \in I} e_i \nu_i(A). \tag{11}$$

The proof of this theorem and all other proofs appear in Appendix B.

Now suppose that there is *probability agreement*, or more exactly that all probabilities can be transformed into agreement. That is, suppose for each $i \in I^+$ there is a legal transform μ_i' of μ_i such that $\mu_i'(A) = \mu_g'(A)$ for all $i \in I$ and all $A \in \mathcal{S}'$. Then for each $i \in I$ take the e_i from Theorem 1 and let $\nu_i' = e_i \nu_i$. Let $\nu_g' = \nu_g$. These are legal transformations. Then

$$U_g'(A) = \frac{\nu_g'(A)}{\mu_g'(A)} = \frac{\nu_g(A)}{\mu_g'(A)} = \textstyle\sum_{i \in I} \dfrac{e_i \nu_i(A)}{\mu_g'(A)}$$

$$= \textstyle\sum_{i \in I} \dfrac{\nu_i'(A)}{\mu_i'(A)} = \sum_{i \in I} U_i'(A).$$

So we have this

Corollary to Theorem 1. *Under Assumptions* 1-4, $\succcurlyeq_g$ *is utilitarian if for each* $i \in I^+$ *there is a legal transform* μ_i' *of* μ_i *such that* $\mu_i'(A) = \mu_g'(A)$ *for all* $i \in I$ *and all* $A \in \mathcal{S}'$.

Probability agreement, then, is a sufficient condition for the Utilitarian Theorem. (Actually, the converse of the Corollary is also true, although that is not proved in this paper. Probability agreement is a necessary and sufficient condition for social preferences to be utilitarian.)

To establish the Utilitarian Theorem, then, we would only need to establish a version of the Probability Agreement Theorem: that individual and social probabilities can be transformed to make them all equal. Assumptions 1–4, however, are not enough to ensure that this is true. The most that can be said is that the individual and social preferences must all agree about the probabilities of propositions that everyone finds indifferent to T. (These probabilities cannot be transformed anyway.) Let $\mathcal{T}$ be $\{A \in \mathcal{S}' : A \approx_i T \text{ for all } i \in I\}$.

Theorem 2. *Under Assumptions* 1–4, $\mu_i(A) = \mu_g(A)$ *for all* $i \in I$ *and all* $A \in \mathcal{T}$.

But outside $\mathcal{T}$ it may not be possible to transform probabilities into agreement. Consequently the Utilitarian Theorem can fail. Examples 1 and 2 in Appendix A are counterexamples. Before we can prove the Utilitarian Theorem we shall need two major new assumptions:

Assumption 5. For all $i \in I$ and all $A \in \mathcal{S}'$ there is a $B \in \mathcal{S}'$ such that

$$B \approx_i A \quad \text{and} \quad B \approx_j T \quad \text{for all } j \in I - \{i\}.$$

Assumption 6. For all $i \in I$, $\succcurlyeq_i$ is unbounded.

Assumption 5 says that there are propositions over the whole range of i's preferences, from top to bottom, that everyone else considers indifferent to T. It says that the different people's preferences are independent of each other in a sense. I shall discuss this independence assumption and Assumption 6 in Section 7. Assumption 4 contains a minimal independence assumption of the same sort. Note that Assumptions 5 and 6 together imply Assumption 4.

Granted assumptions 5 and 6, the Probability Agreement Theorem can be proved:

Theorem 3. *Under Assumption* 1, 2, 3, 5 *and* 6, $\mu_i(A) = \mu_g(A)$ for all $i \in I$ and all $A \in \mathcal{S}'$.

(This says probabilities actually agree, not merely that they can be transformed into agreement, because under Assumption 6 probabilities cannot be transformed.) From Theorem 3 and the Corollary to Theorem 1, the Utilitarian Theorem follows immediately:

Theorem 4. *Under Assumptions* 1, 2, 3, 5 *and* 6, $\succcurlyeq_g$ *is utilitarian.*

7. CONCLUSIONS

The Utilitarian Theorem, as I described it in Section 1, says that if individual and social preferences are coherent (Assumptions 1 and 2), and if social preferences are Paretian (Assumption 3), then it follows that social preferences are utilitarian. Theorem 4 tells us that Assumptions 5 and 6 are sufficient for the truth of this theorem.

Neither would be sufficient on its own. This is shown by the counterexamples contained in Appendix A. Example 1 satisfies Assumption 5 and Example 2 satisfies Assumption 6, and both examples satisfy Assumptions 1, 2 and 3. But in neither example are social preference utilitarian.

Nevertheless, neither Assumption 5 nor Assumption 6 is a necessary condition for social preferences to be utilitarian, even given Assumptions 1, 2 and 3. There are therefore sufficient conditions that are weaker than Assumptions 5 and 6 together. Indeed, we already know some from the Corollary to Theorem 1: the condition that all probabilities are in agreement (or can be transformed into agreement) and the very weak Assumption 4 are together sufficient for social preferences to be utilitarian, given Assumptions 1, 2 and 3. And probability agreement is, in a sense, the weakest possible condition, because it is actually necessary. But, if we are interested in finding weaker sufficient conditions than Assumptions 5 and 6, probability agreement is not what we are looking for. To assume agreement about probabilities from the start would be to give up the generality we gained by allowing for subjective probabilities in the first place. The Probability Agreement Theorem, I explained in Section 2, is important in its own right. What we are looking for are conditions that, together with Assumptions 1, 2 and 3, are sufficient to force probabilities into agreement: sufficient conditions for the Probability Agreement Theorem, that is. These conditions will then be sufficient for the Utilitarian Theorem too.

Ideally, one might hope for conditions that are so weak as to be actually necessary (given Assumptions 1, 2 and 3) for probability agreement. But if these are to be, like Assumptions 5 and 6, conditions on individual preferences only, this hope is a vain one. There are no conditions on individual preferences that are necessary for probability agreement, except for conditions that are *trivially* necessary. (By a "trivially necessary" condition, I mean one that follows from probability agreement itself and nothing else. For example: that any two people agree about probabilities.) People's probabilities could just *happen* to agree with social probabilities, whatever their preferences might be like in other respects. To see this, suppose we had a condition P on individual preferences that, given Assumptions 1, 2 and 3, was non-trivially necessary for probability agreement. Let A be probability agreement; more exactly, let A be the proposition that individual and social probabilities can be transformed into agreement. Then 1, 2, 3 and A together imply P. But P must not be implied by A alone; that would make it trivially necessary. So it must be possible for A to be true and P false. Take a situation where A is true. Assumptions 1 and 2 must also be true, because without them (coherence of the individual and social preferences) A does not even make sense: individual and social probabilities are not defined. For each person, pick probabilities that agree with social probabilities, and, using these probabilities, pick a utility function that represents her preferences. Let the "cooked-up social utility function" be the sum of all these functions. This function will define "cooked-up social preferences" that are coherent and Paretian. The social preferences in the given situation may be different from these. But imagine a situation that has the same individual preferences, and where the social preferences *are* these cooked-up ones. In this imaginary situation, conditions 1, 2, 3 and A would all be true. P would therefore be true because these conditions imply P. But P is a condition on individual preferences, which are the same in the imaginary situation as in the given one. So P is true in the given situation too. Whenever A is true, therefore, P is true. And that contradicts what we assumed.

There are, certainly, conditions involving *social* preferences that are non-trivially necessary and non-trivially sufficient for probability agreement, given Assumptions 1, 2 and 3. Here is one: either social preferences are Paretian and there is probability

agreement, or social preferences are not Paretian and there is not probability agreement. But I take it that adding further condition on social preferences would not leave us with anything recognizable as the Utilitarian Theorem. The idea of this theorem, I take it, is to derive the remarkable conclusion that social preferences are utilitarian from just those two particular conditions on social preferences: that they are coherent and Paretian.

So there is not much point in looking for necessary conditions. One might find sufficient conditions that are weaker than Assumptions 5 and 6. But my own belief, born out of work on the proof and counterexamples, is that no weaker conditions would be simple enough and general enough to be interesting. For the moment, at least, our trust in the Utilitarian Theorem will have to be based on the conditions we have available: Assumptions 5 and 6. How plausible are they?

Assumption 5 is very demanding. The values people assign to prospects are likely to be correlated to some extent. A nuclear war, for instance, comes near the bottom of most people's scales of preference. But Assumption 5 requires there to be some prospect that is just as bad as this for one person, but that other people regard with equanimity. This is implausible. I doubt the Utilitarian Theorem could be proved on the basis of a much weaker independence assumption than this. My proof depends crucially on the existence of propsects that are very good or very bad for one person and neutral for everyone else. So I doubt if the objectionable feature of Assumption 5 could be eliminated.

Other versions of the Utilitarian Theorem, within other versions of decision theory, also rely on independence assumptions of the same sort. Harsanyi's (1955) original version took probabilities to be objective. A more rigorous proof with objective probabilities has been published by Peter Fishburn (1984). Fishburn assumes that for each person i there are prospects A_i and B_i such that $A_i \succ_i B_i$ and $A_i \approx_j B_i$ for all j other than i. This is a very weak independence assumption. Indeed, it is equivalent to my Assumption 4. My Corollary to Theorem 1, which assumes probability agreement, corresponds closely to Fishburn's theorem. It, too, requires Assumption 4.

Peter Hammond (1983) assumes implicitly that the range of the vector $\langle u_{11}, u_{12}, \ldots, u_{21}, \ldots u_{hn}\rangle$, where u_{ir} is person i's utility in state of nature r, is a product set ΠS_{ir}, where S_{ir} is the range of u_{ir}. This means that one person's utility in one state can move over the whole of its range whilst the utilities for every other person-state pair remain constant. This is an extremely strong independence assumption. Other proofs use even stronger ones. Robert Deschamps and Louis Gevers (1979) assume that the range of $\langle u_{11}, u_{12}, \ldots, u_{21}, \ldots u_{hn}\rangle$ is the whole of Euclidean space. Hammond's earlier proof (1981) helps itself to variations in u_{ir} without any constraint. However, there is reason to think that strong independence assumptions can be dispensed with. In his 1983 proof Hammond required the range of utilities to be a product set because he made use of a theorem of W. M. Gorman's (1968) that assumes a range of this shape. But Gorman tells me his theorem could be proved on the basis of a much weaker assumption about the range. So Hammond's proof could actually be founded on a much weaker independence axiom.

It seems, then, that the need for a very strong independence assumption in proving the Utilitarian Theorem may be confined to Bolker-Jeffrey expected utility theory.

The need for Assumption 6, the unboundedness assumption, is also confined to this theory, because it makes no sense in other theories. But one might also say that an assumption with the same role is already implicit in other theories. The effect of Assumption 6 is to cancel out a sort of extra generality that the Bolker-Jeffrey theory initially possesses beyond other expected utility theories: the wider range of utility transformations it allows. When preferences are unbounded, only linear transformations are possible, as

in other theories. I did not expect this conclusion. I expected that having a wide range of utility functions available for representing preferences would actually make it easier to find a social utility function that is the sum of individual functions. But the opposite is true. This suggests to me that a tight determination of utilities may be playing a part in making the Utilitarian Theorem work in other versions of expected utility theory too. As I explained in Section 4, this tight determination is the result of supposing that a person has preferences over arbitrarily constructed gambles, including gambles that are causally impossible. It may be, then, that social preferences are forced to be utilitarian only because we insist that they should be coherent and Paretian over gambles that are causally impossible. This is some reason to be cautious about the theorem.

It is not easy to assess the plausibility of Assumption 6 itself: that everyone's preferences are unbounded. I said in Section 4 that Bolker and Jeffrey have not made it clear just what unbounded preferences are like. So it is hard to know whether to expect them to be common or uncommon. But to assume that everyone's preferences are unbounded does seem, at least, to be asking for a large coincidence.

I conclude, then, that the Utilitarian Theorem, though it can be proved within the Bolker–Jeffrey theory, needs to be treated with caution.

In another article (Broome (1987)) I have reinterpreted the theorem in terms of a person's *good*, rather than her preferences. A person's utility I take to represent her "betterness relation"

$$\_ \text{ is at least as good for the person as } \_$$

instead of her preference relation

$$\_ \text{ is preferred or indifferent to } \_.$$

But, as that article explains, the notion of betterness applied to uncertain prospects presupposes probabilities that are the same for everyone. The reinterpreted theorem therefore does not require Assumptions 5 and 6. It is shown to be true by the Corollary to Theorem 1. This requires only Assumption 4, the weakest of independence assumptions. I therefore have faith in the reinterpreted theorem.

APPENDIX A: COUNTEREXAMPLES

The examples below satisfy all the conditions of Theorem 4 except, for Example 1, Assumption 5 and, for Example 2, Assumption 6. But in both examples the theorem fails. I need to set up some preliminaries before coming to the examples.

The examples need to consist of preferences defined on a complete atomless Boolean algebra. The algebra I shall use—unfortunately no simpler one exists—is defined as follows. Let Ω be the unit interval $[0, 1]$ and let λ be Lebesgue measure on Ω. Take the set of Lebesgue-measurable subsets of Ω, and let $\mathscr{L}$ be the partition of this set into equivalence classes of sets that differ only by a set of measure zero. Then $\mathscr{L}$ is a complete atomless Boolean algebra.

A complete Boolean algebra $\mathscr{S}$ on which a strictly positive measure (a measure μ such that $\mu(A) > 0$ for all A in $\mathscr{S}$ other than F) exists is called a *measure algebra* (see Sikorski (1960, p. 149)). The algebra described above is the *measure algebra of* λ.

There is a way of showing legal transformations geometrically that will be useful (see Bolker (1966 pp. 304–306)). Let N be the vector space of measures on $\mathscr{S}$. In N the non-negative measures form a convex cone M. Let $\bar{M}$ be the convex cone of non-positive measures, and $\hat{N}$ the set of strictly signed measures (measures that are neither non-negative nor non-positive). Since μ is a probability measure, it is in M and is distinct from 0. If $\nu = 0$, then $U(A) = 0$ for all A in $\mathscr{S}'$, and all propositions are indifferent. Let us assume this is not so. Then there is an A in $\mathscr{S}'$ such that $\nu(A) \neq 0$. But $\nu(A) + \nu(\sim A) = \nu(A \vee \sim A) = \nu(T) = 0$. So $\nu(A) = -\nu(\sim A)$. Therefore ν is a strictly signed measure. The measures μ and ν consequently form a basis for a two-dimensional subspace of N. And μ' and ν' given by (7) form another basis for the same subspace. This subspace is shown in Figure 1. All legal transforms lie on the dashed lines.

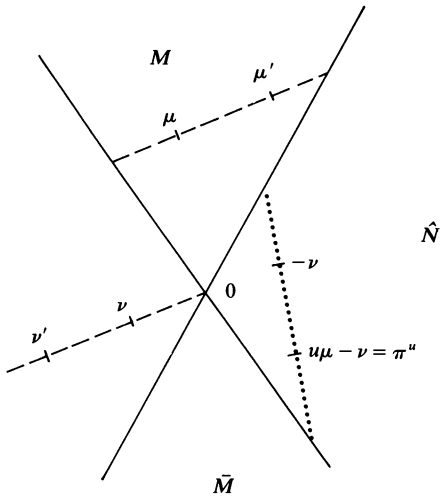

FIGURE 1

Consider the measures $\pi^u = u\mu - \nu$ for all u in the range of U on $\mathscr{S}'$. We shall see that what makes these measures useful is this: for any A in $\mathscr{S}'$, if $u = U(A)$ then $\pi^u(A) = 0$. This is true because $\pi^u(A) = u\mu(A) - \nu(A) = \mu(A)\,[u - \nu(A)/\mu(A)] = \mu(A)[u - U(A)] = 0$. In Figure 1 these measures lie along the dotted line. The next paragraph proves that the line extends from the frontier of $\bar{M}$ to the frontier of M but no further. (It may or may not include the frontier points, depending on whether or not U attains its bounds.)

First, $\pi^u = u\mu - \nu$ is a strictly positive or a strictly negative measure if and only if u is outside the range of U on $\mathscr{S}'$. For if π^u is strictly positive or strictly negative there is no $A \in \mathscr{S}'$ such that $\pi^u(A) = 0$, no A, that is to say, such that $U(A) = u$. And if π^u is neither strictly positive nor strictly negative, there are A and B in $\mathscr{S}'$ such that $\pi^u(A) \geqq 0$ and $\pi^u(B) \leqq 0$. Then $U(A) \leqq u$ and $U(B) \geqq u$. But the range of U on $\mathscr{S}'$ is an interval (Bolker (1966, p. 294, Lemma 1.12)), so there is a $C \in \mathscr{S}'$ such that $U(C) = u$. Next, if there is a $\hat{u} < u$ such that $\pi^{\hat{u}}$ is a non-negative measure then π^u is a strictly positive measure. For any $A \in \mathscr{S}'$ has $\pi^{\hat{u}}(A) \geqq 0$, and so $\pi^u(A) = \pi^{\hat{u}}(A) + (u - \hat{u})\mu(A) > 0$ because $\mu(A) > 0$. Similarly, if there is a $\hat{u} > u$ such that $\pi^{\hat{u}}$ is a non-positive measure then π^u is a strictly negative measure.

If U is unbounded above and below, the plane spanned by μ and ν meets M and $\bar{M}$ in a single line. It is shown in Figure 2.

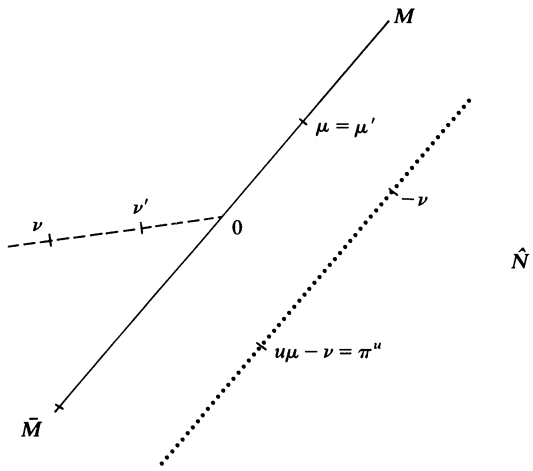

FIGURE 2

Example 1. Let Ω be the unit interval $[0, 1]$, let λ be Lebesgue measure on Ω, let $\mathscr{L}$ be the measure algebra of λ, and let $\mathscr{L}' = \mathscr{L} - \{F\}$. Let there be two people. Let the various measures be as follows. For all $A \in \mathscr{L}$,

$$\nu_1(A) = \int_A f_1(x)d\lambda(x) \quad \text{where } f_1(x) = -1 \quad \text{for } 0 \leqq x < \tfrac{1}{2}$$

$$= 1 \quad \text{for } \tfrac{1}{2} \leqq x \leqq 1$$

and

$$\nu_2(A) = \int_A f_2(x)d\lambda(x) \quad \text{where } f_2(x) = -1 \quad \text{for } 0 \leqq x < \tfrac{1}{4} \text{ and } \tfrac{3}{4} \leqq x \leqq 1$$

$$= 1 \quad \text{for } \tfrac{1}{4} \leqq x < \tfrac{3}{4}.$$

$$\nu_g = \tfrac{1}{2}(\nu_1 + \nu_2),$$

$$\mu_1 = \mu_2 = \lambda,$$

$$\mu_g = \lambda + \tfrac{1}{4}(\nu_1 - \nu_2).$$

Consider the measures $\pi_1^{u_1} = u_1\mu_1 - \nu_1$, $\pi_2^{u_2} = u_2\mu_2 - \nu_2$ and $\pi_g^{u_g} = u_g\mu_g - \nu_g$, where u_1, u_2 and u_g are respectively members of the ranges of U_1, U_2 and U_g on $\mathscr{L}'$. (These ranges are all the interval $[-1, 1]$.) All these measures are contained in a single two-dimensional plane in the vector space of measures of $\mathscr{L}$. This plane is shown in Figure 3.

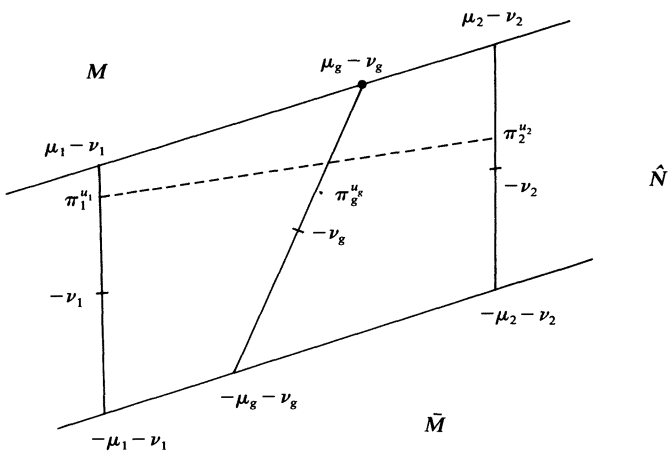

FIGURE 3

I shall show first that the social preferences in this example are Paretian. Take A and B in $\mathscr{L}'$ such that $A \approx_1 B$ and $A \approx_2 B$. Let $u_1 = U_1(A) = U_1(B)$ and let $u_2 = U_2(A) = U_2(B)$. Then $\pi_1^{u_1}(A) = \pi_1^{u_1}(B) = \pi_2^{u_2}(A) = \pi_2^{u_2}(B) = 0$. The measures $\pi_1^{u_1}$ and $\pi_2^{u_2}$ are shown in Figure 3. Whatever the values of u_1 and u_2 within the range of U_1 and U_2, there is a u_g (specifically $2(u_1 + u_2)/(4 + u_1 - u_2)$) within the range of U_g such that $\pi_g^{u_g}$ is a convex combination of $\pi_1^{u_1}$ and $\pi_2^{u_2}$. For this u_g, $\pi_g^{u_g}(A) = 0 = \pi_g^{u_g}(B)$. So $U_g(A) = u_g = U_g(B)$. That is to say, $A \approx_g B$. This shows that condition (9) of the Paretian assumption (Section 6) is satisfied. Furthermore u_g plainly increases with u_1 and u_2. This shows that condition (10) is satisfied.

And next I shall show that the social preferences are not utilitarian. I explained that all the legal transforms of μ_1 lie in the subspace spanned by μ_1 and ν_1. All the legal transforms of μ_2 and μ_g lie, respectively, in the subspaces spanned by μ_2 and ν_2, and by μ_g and ν_g. But these three subspaces have no non-zero vector in common. So the probabilities cannot be transformed to bring them into agreement. Consequently the social preferences are not utilitarian. But since I have not proved that agreement in probabilities is a necessary

condition for social preferences to be utilitarian, I need to give a further proof of this. If the social preferences were utilitarian there would be legal transforms U_1', U_2' and U_g' of U_1, U_2 and U_g such that $U_g' = U_1' + U_2'$. That is to say there would be coefficients a_1, a_2, a_g, c_1, c_2, and c_g, all satisfying conditions (8), such that

$$\frac{a_g U_g(A)}{c_g U_g(A)+1} = \frac{a_1 U_1(A)}{c_1 U_1(A)+1} + \frac{a_2 U_2(A)}{c_2 U_2(A)+1} \quad \text{for all } A \in \mathscr{L}'. \tag{12}$$

Tabulated below are the utilities of six intervals in $\mathscr{L}'$.

	U_1	U_2	U_g
$[0, \frac{1}{4}]$	-1	-1	-1
$[\frac{1}{4}, \frac{1}{2}]$	-1	1	0
$[\frac{1}{2}, \frac{3}{4}]$	1	1	1
$[\frac{3}{4}, 1]$	1	-1	0
$[0, \frac{1}{2}]$	-1	0	$-\frac{2}{3}$
$[\frac{1}{2}, 1]$	1	0	$\frac{2}{5}$

Substituting these values into equation (12) gives six equations in the five unknowns a_1/a_g, a_2/a_g, c_1, c_2 and c_g. A little algebra shows that no values for the unknowns can satisfy all six equations at once.

Finally I shall show that the example satisfies Assumption 5. This assumption requires that for any u_1 in the range of U_1 there is a $B \in \mathscr{L}'$ with $U_1(B) = u_1$ and $U_2(B) = 0$. Consider the vector-valued measure $\langle \mu_1, \nu_1, \nu_2 \rangle$. For $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$ respectively, this takes on the values $\langle \frac{1}{2}, -\frac{1}{2}, 0 \rangle$ and $\langle \frac{1}{2}, \frac{1}{2}, 0 \rangle$. By Liapounoff's Theorem (see Appendix B) there is some $B \in \mathscr{L}'$ for which it takes on the value $\langle \frac{1}{2}, \frac{1}{2}u_1, 0 \rangle$ for any u_1 in the range $-1 \leqq u_1 \leqq 1$, which is the range of U_1. This B has $U_1(B) = u_1$ and $U_2(B) = 0$.

This example shows that coherent Paretian social preferences need not be utilitarian. This refutes the Utilitarian Theorem, and undermines whatever support it may give to utilitarianism. The theorem claims that coherent Paretian social preferences just have to be utilitarian. However, although the given social preferences are not utilitarian, it is possible to find some that are. Simply change μ_g to λ and leave ν_g unchanged. So in this example utilitarianism is possible. In Example 2, on the other hand, there are coherent Paretian social preferences, but no utilitarian social preferences are possible.

Example 2. Let $\mathscr{L}$ and λ be as in Example 2. Define the measure ρ by

$$\rho(A) = \int_A [\log(x) - \log(1-x)] d\lambda(x) \quad \text{for all } A \in \mathscr{L}.$$

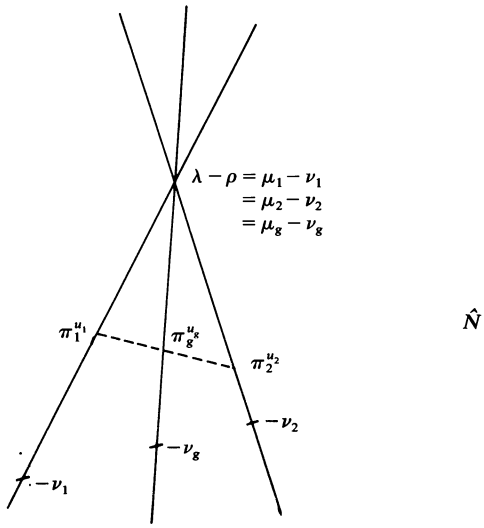

FIGURE 4

Let

$$\mu_1(A) = \int_A (\tfrac{1}{2} + x)\, d\lambda(x) \qquad \text{for all } A \in \mathscr{L},$$

$$\mu_2(A) = \int_A (\tfrac{3}{2} - x)\, d\lambda(x) \qquad \text{for all } A \in \mathscr{L},$$

$$\mu_g = \lambda,$$

$$\nu_1 = \rho - \lambda + \mu_1,$$

$$\nu_2 = \rho - \lambda + \mu_2,$$

$$\nu_g = \rho.$$

The measures $\pi_1^{u_1} = u_1\mu_1 - \nu_1$, $\pi_2^{u_2} = u_2\mu_2 - \nu_2$ and $\pi_g^{u_g} = u_g\mu_g - \nu_g$ are once again confined to a single plane. It is shown in Figure 4. This time the ranges of U_1, U_2 and U_g are unbounded above and below, because ρ is unbounded above and below. So the example satisfies Assumption 6.

Take any A in $\mathscr{L}'$. Let $u_1 = U_1(A) = 1 + [\rho(A) - \lambda(A)]/\mu_1(A)$, and $u_2 = U_2(A) = 1 + [\rho(A) - \lambda(A)]/\mu_2(A)$. Since $\mu_1(A) > 0$ and $\mu_2(A) > 0$, $(u_1 - 1)$ and $(u_2 - 1)$ have the same sign. In Figure 4, $\pi_1^{u_1}$ and $\pi_2^{u_2}$ are both above, or both below, or both at $(\lambda - \rho)$. The same argument as I used for Example 1 shows that social preferences are Paretian. But they are not utilitarian because only linear transformations of utilities are legal, and there are no linear transformations that make U_g the sum of U_1 and U_2.

APPENDIX B: PROOFS

The proofs in this appendix depend on a theorem of Liapounoff's, which is proved by Bolker (1966, p. 295):

Liapounoff's Theorem. *The range of a vector-valued measure on an atomless measure algebra is convex.*

A measure algebra is a complete Boolean algebra on which a strictly positive measure exists. Bolker's Existence Theorem shows that, under the conditions of that theorem, $\mathscr{S}$ is an atomless measure algebra. So Liapounoff's Theorem applies to $\mathscr{S}$,

Proof of Theorem 2. Let γ be the vector-valued measure $\langle \mu_1, \ldots, \mu_h, \mu_g, \nu_1, \ldots, \nu_h, \nu_g \rangle$ and let Γ be its range on $\mathscr{S}$. Γ contains 0 and also $\gamma(T) = \langle \mu_1(T), \ldots, \mu_h(T), \mu_g(T), \nu_1(T), \ldots, \nu_h(T), \nu_g(T) \rangle = \langle 1, \ldots, 1, 1, 0 \ldots 0, 0 \rangle$. So because Γ is convex by Liapounoff's Theorem, it contains $\langle p, \ldots p, p, 0, \ldots, 0, 0 \rangle$ for any real p with $0 \leqq p \leqq 1$.

By Assumption 3, if $A \approx_i T$ for all $i \in I$, then $A \approx_g T$. So for any $A \in \mathscr{S}'$, A is in $\mathscr{T}$ if and only if $\nu_i(A) = 0$ for all $i \in I^+$. I shall show that if A is in $\mathscr{T}$ and $\mu_g(A) = p$ then $\mu_i(A) = p$ for all $i \in I$. Suppose for some $i \in I$ $\mu_i(A) \neq p$. Let A_i be as defined in Assumption 4, so that $\nu_i(A_i) \neq 0$ but $\nu_j(A_i) = 0$ for all $j \in I - \{i\}$. I have shown in the previous paragraph that there is a B in $\mathscr{T}$ such that $\mu_i(B) = \mu_g(B) = p$. By a lemma of Bolker's (1966, p. 300, Lemma 3.1) there are mutally contrary elements A', A_i', B' of $\mathscr{S}'$ such that $\gamma(A') = \gamma(A)/8$, $\gamma(A_i') = \gamma(A_i)/8$ and $\gamma(B') = \gamma(B)/8$. In particular $\nu_j(A') = \nu_j(B') = 0$ for all $j \in I^+$, $\nu_j(A_i') = 0$ for all $j \in I - \{i\}$, but $\nu_i(A_i') \neq 0$. Also $\mu_g(A') = \mu_g(B') = p/8$, and $\mu_i(B') = p/8$ but $\mu_i(A') \neq p/8$. Consider the disjunctions of contraries $(A_i' \vee A')$ and $(A_i' \vee B')$. For all $j \in I - \{i\}$, $(A_i' \vee A') \approx_j (A_i' \vee B')$ because

$$\nu_j(A_i' \vee A') = \nu_j(A_i') + \nu_j(A') = 0 = \nu_j(A_i') + \nu_j(B') = \nu_j(A_i' \vee B').$$

Also $(A_i' \vee A') \approx_g (A_i' \vee B')$ because

$$U_g(A_i' \vee A') = \frac{\nu_g(A_i') + \nu_g(A')}{\mu_g(A_i') + \mu_g(A')} = \frac{\nu_g(A_i') + \nu_g(B')}{\mu_g(A_i') + \mu_g(B')} = U_g(A_i' \vee B').$$

But it is not true that $(A_i' \vee A') \approx_i (A_i' \vee B')$ because

$$U_i(A_i' \vee A') = \frac{\nu_i(A_i') + \nu_i(A')}{\mu_i(A_i') + \mu_i(A')} \neq \frac{\nu_i(A_i') + \nu_i(B')}{\mu_i(A_i') + \mu_i(B')} = U_i(A_i' \vee B').$$

And this contradicts Assumption 3. ‖

The proof of Theorem 3 is based on the following Lemma. The Lemma requires only Assumptions 1-4, and it also supplies a proof for Theorem 1. (A shorter proof of Theorem 1 is available, but not given here.)

Lemma. *Let γ be the vector-valued measure $\langle \mu_1, \ldots, \mu_h, \mu_g, \nu_1, \ldots, \nu_h, \nu_g \rangle$, taking values in R^{2h+2}. Let Γ be its range on $\mathscr{S}$. Let H be the subspace of R^{2h+2} defined by $\nu_i = 0$ for all $i \in I$. Let L be the one-dimensional*

subspace spanned by $\gamma(T) = \langle 1, \ldots, 1, 1, 0, \ldots, 0, 0\rangle$. *Under Assumptions* 1-4, Γ *is contained in an* $h+1$ *dimensional subspace* Γ^* *of* R^{2h+2}, *and* $H \cap \Gamma^* = L$.

Proof. Take the A_i of the Assumption 4. Then

$$\gamma(A_1) = \langle \mu_1(A_1), \ldots, \mu_h(A_1), \mu_g(A_1), \nu_1(A_1), 0, \ldots, 0, \nu_g(A_1)\rangle$$

$$\gamma(A_2) = \langle \mu_1(A_2), \ldots, \mu_h(A_2), \mu_g(A_2), 0, \nu_2(A_2), \ldots, 0, \nu_g(A_2)\rangle$$

$$\gamma(A_h) = \langle \mu_1(A_h), \ldots, \mu_h(A_h), \mu_g(A_h), 0, 0, \ldots, \nu_h(A_h), \nu_g(A_h)\rangle$$

and

$$\gamma(T) = \langle 1, \ldots, 1, 1, 0, 0, \ldots, 0, 0\rangle.$$

For all $i \in I$, $\nu_i(A_i) \neq 0$. So these vectors are linearly independent, and they form a basis for an $h+1$ dimensional subspace. No non-zero linear combination of the $\gamma(A_i)$ is contained in H, so the subspace meets H only in the line L spanned by $\gamma(T)$. I shall show that for any $A \in \mathcal{S}$, $\gamma(A)$ is a linear combination of $\gamma(T)$ and the $\gamma(A_i)$. That will complete the proof.

Take any $A \in \mathcal{S}$. For all $i \in I$, if $\nu_i(A)$ is zero or has the opposite sign to $\nu_i(A_i)$, define A_i' as A_i. If $\nu_i(A)$ has the same sign as $\nu_i(A_i)$, define A_i' as $\sim A_i$. I shall show that $\gamma(A)$ is a linear combination of $\gamma(T)$ and the $\gamma(A_i')$. Since for all $i \in I$ either $\gamma(A_i') = \gamma(A_i)$ or

$$\begin{aligned}\gamma(A_i') &= \gamma(\sim A_i)\\ &= \langle 1 - \mu_1(A_i), \ldots, 1 - \mu_h(A_i), 1 - \mu_g(A_i), 0, \ldots, -\nu_i(A_i), \ldots, 0, -\nu_g(A_t)\rangle\\ &= \gamma(T) - \gamma(A_i),\end{aligned}$$

it will follow, as required, that $\gamma(A)$ is a linear combination of $\gamma(T)$ and the $\gamma(A_i)$.

For all $i \in I$ define

$$f_i = -\frac{\nu_i(A)}{\nu_i(A_i')}, \tag{13}$$

so $f_i \geqq 0$. Define f as $\sum_{i \in I} f_i$. Define

$$\hat{\gamma} = \frac{1}{1+f}[\gamma(A) + \textstyle\sum_{i \in I} f_i \gamma(A_i')].$$

Then $\hat{\gamma}$ is a convex combination of $\gamma(A)$ and the $\gamma(A_i')$. Since all these vectors belong to Γ, which is convex by Liapounoff's Theorem, $\hat{\gamma}$ belongs to Γ. So there is a $C \in \mathcal{S}$ such that $\gamma(C) = \hat{\gamma}$. For all $j \in I$,

$$\begin{aligned}\nu_j(C) &= \frac{1}{1+f}[\nu_j(A) + \textstyle\sum_{i \in I} f_i \nu_j(A_i')]\\ &= \frac{1}{1+f}[\nu_j(A) + f_j \nu_j(A_j')] \quad \text{because } \nu_j(A_i') = 0 \text{ for all } i \in I - \{j\}\\ &= 0 \qquad\qquad\qquad\qquad \text{by (13).}\end{aligned}$$

So for all $j \in I$, $C \approx_j T$. It follows by Theorem 2 that $\mu_j(C) = \mu_g(C)$ for all $j \in I$ and by Assumption 3 that $\nu_g(C) = 0$. So

$$\mu_j(A) + \textstyle\sum_{i \in I} f_i \mu_j(A_i') = \mu_j(C)(1+f) = \mu_g(C)(1+f) = \mu_g(A) + \sum_{i \in I} f_i \mu_g(A_i') \quad \text{for all } j \in I$$

and

$$\nu_g(A) + \textstyle\sum_{i \in I} f_i \nu_g(A_i') = \nu_g(C)(1+f) = 0.$$

Consequently,

$$\mu_j(A) = -\textstyle\sum_{i \in I} f_i \mu_j(A_i') + [\mu_g(A) + \sum_{i \in I} f_i \mu_g(A_i')] \quad \text{for all } j \in I \tag{14}$$

and

$$\nu_g(A) = -\textstyle\sum_{i \in I} f_i \nu_g(A_i'). \tag{15}$$

We know already that

$$\nu_j(A) = -\sum_{i\in I} f_i \nu_j(A_i') \quad \text{for all } j \in I, \tag{16}$$

and of course

$$\mu_g(A) = -\sum_{i\in I} f_i \mu_g(A_i') + [\mu_g(A) + \sum_{i\in I} f_i \mu_g(A_i')]. \tag{17}$$

Equations (14), (15), (16) and (17) amount to:

$$\gamma(A) = -\sum_{i\in I} f_i \gamma(A_i') + [\mu_g(A) + \sum_{i\in I} f_i \mu_g(A_i')]\gamma(T).$$

That is to say, $\gamma(A)$ is a linear combination of $\gamma(T)$ and the $\gamma(A_i')$. ‖

Proof of Theorem 1. From (13) and (15),

$$\nu_g(A) = \sum_{i\in I} \frac{\nu_g(A_i')}{\nu_i(A_i')} \nu_i(A).$$

Let $e_i = \nu_g(A_i')/\nu_i(A_i') = \nu_g(A_i)/\nu_i(A_i)$. Assumption 3 ensures that $\nu_g(A_i)/\nu_i(A_i) > 0$. ‖

Proof of Theorem 4. For each number $n = 1, 2, 3, \ldots$, take an A^n and a B^n such that $U_1(A^n) > n$, $U_1(B^n) < -n$ and $U_i(A^n) = U_i(B^n) = 0$ for all $i \in I - \{1\}$. Assumptions 5 and 6 ensure that such A^n and B^n exist. Since $\gamma(A^n)$ and $\gamma(B^n)$ are in Γ, the vectors

$$\bar{\gamma}^n = \frac{\gamma(A^n)}{\nu_1(A^n)} = \left\langle \frac{1}{U_1(A^n)}, \frac{\mu_2(A^n)}{\nu_1(A^n)}, \ldots, \frac{\mu_h(A^n)}{\nu_1(A^n)}, \frac{\mu_g(A^n)}{\nu_1(A^n)}, 1, 0, \ldots, 0, e_1 \right\rangle$$

and

$$\underline{\gamma}^n = \frac{\gamma(B^n)}{\nu_1(B^n)} = \left\langle \frac{1}{U_1(B^n)}, \frac{\mu_2(B^n)}{\nu_1(B^n)}, \ldots, \frac{\mu_h(B^n)}{\nu_1(B^n)}, \frac{\mu_g(B^n)}{\nu_1(B^n)}, 1, 0, \ldots, 0, e_1 \right\rangle$$

(where e_1 is as defined in Theorem 1) are both in Γ^*. Therefore $\bar{\gamma}^n - \underline{\gamma}^n$ is in Γ^*. But $\bar{\gamma}^n - \underline{\gamma}^n$ is in the subspace H. Since $\Gamma^* \cap H = L$ by the Lemma, $\bar{\gamma}^n - \underline{\gamma}^n$ is in L. This means that all the first $h+1$ components of $\bar{\gamma}^n - \underline{\gamma}^n$ are equal to each other. The first component is

$$\frac{1}{U_1(A^n)} - \frac{1}{U_1(B^n)} < \frac{2}{n}.$$

So

$$\frac{\mu_i(A^n)}{\nu_1(A^n)} - \frac{\mu_i(B^n)}{\nu_1(B^n)} < \frac{2}{n} \quad \text{for all } i \in I^+ - \{1\}.$$

Since $\nu_1(A^n) > 0$ and $\nu_1(B^n) < 0$ if follows that

$$0 \leqq \frac{\mu_i(A^n)}{\nu_1(A^n)} < \frac{2}{n} \quad \text{for all } i \in I^+ - \{1\}.$$

Since, also,

$$0 \leqq \frac{1}{U_1(A^n)} < \frac{1}{n},$$

the sequence $\{\bar{\gamma}^n\}$ has a limit, which is $\langle 0, 0, \ldots, 0, 0, 1, 0, \ldots, 0, e_1\rangle$. Because Γ^* is closed and $\bar{\gamma}^n \in \Gamma^*$ for all n, this limit is in Γ^*. Call it γ_1. Similarly γ_i is in Γ^* for all $i \in I$, where

$$\gamma_1 = \langle 0, 0, \ldots, 0, 0, 1, 0, \ldots, 0, e_1\rangle$$

$$\gamma_2 = \langle 0, 0, \ldots, 0, 0, 0, 1, \ldots, 0, e_2\rangle$$

and

$$\cdots$$

$$\gamma_h = \langle 0, 0, \ldots, 0, 0, 0, 0, \ldots, 1, e_h\rangle.$$

Also

$$\gamma(T) = \langle 1, 1, \ldots, 1, 1, 0, 0, \ldots, 0, 0\rangle.$$

These $h+1$ vectors are linearly independent. So they form a basis for Γ^*. Consequently, for any A, $\gamma(A)$ is a linear combination of them. So the first $h+1$ components of $\gamma(A)$ must be equal to each other. ‖

Acknowledgement. I am grateful for the helpful comments I have received from W. M. Gorman, Richard Jeffrey, David Kreps, Adam Morton and the editor and referees. And I am grateful to Mary Harthan for the accurate typing of the mathematics. I did part of the work on this paper while I was a visitor at Princeton University.

REFERENCES

BALCH, M., McFADDEN, D. and WU, S. (eds) (1974) *Essays on Economic Behavior Under Uncertainty* (Amsterdam: North-Holland).

BARRO, R. J. and GORDON, D. B. (1983), "Rules, discretion and reputation in a model of monetary policy", *Journal of Monetary Economics*, **12**, 101–121.

BOLKER, E. D. (1966), "Functions resembling quotients of measures", *Transactions of the American Mathematical Society*, **124**, 292–312.

BOLKER, E. D. (1967), "A simultaneous axiomatization of utility and subjective probability", *Philosophy of Science*, **34**, 333–340.

BOLKER, E. D. (1974), "Remarks on 'subjective expected utility for conditional primitives'", in Balch, McFadden and Wu (1974), 79–82.

BORDER, K. C. (1985), "More on Harsanyi's cardinal welfare theorem", *Social Choice and Welfare*, **2**, 279–281.

BROOME, J. (1982), "Uncertainty in welfare economics, and the value of life", in Jones-Lee (1982), 201–216.

BROOME, J. (1987), "Utilitarianism and expected utility", *Journal of Philosphy*, **84**, 405–422.

BROOME, J. (1989), "Should social preferences be consistent?", *Economics and Philosophy*, **5**, 7–17.

BROOME, J. (1990) *Weighing Goods* (Oxford: Basil Blackwell).

BROOME, J. (1991) "Utilitarian metaphysics?", in Elster and Roemer (1991).

BUTTS, R. and HINTIKKA, J. (eds) (1977) *Foundational Problems in the Special Sciences* (Dordrecht: Reidel).

DESCHAMPS, R. and GEVERS, L. (1979), "Separability, risk-bearing and social welfare judgements", in Laffont (1979), 145–160.

EELLS, E. (1982) *Rational Decision and Causality* (Cambridge: Cambridge University Press).

ELSTER, J. and ROEMER, J. (eds) (1991) *Interpersonal Comparisons of Well-being* (Cambridge: Cambridge University Press).

FEIWEL, G. R. (ed) (1987) *Arrow and the Foundations of the Theory of Economic Policy* (London: Macmillan).

FENTON, J. H. (1988) *Probability and Causality* (Dordrecht: Reidel).

FISHBURN, P. C. (1984), "On Harsanyi's utilitarian cardinal welfare theorem", *Theory and Decision*, **17**, 21–28.

FRYDMAN, R., O'DRISCOLL, G. P. and SCHOTTER, A. (1982), "Rational expectations of government policy: an application of Newcomb's problem", *Southern Economic Journal*, **49**, 311–319.

GIBBARD, A. and HARPER, W. L. (1978), "Counterfactuals and two kinds of expected utility", in Hooker, Leach and McClennen (1978), 125–162.

GORMAN, W. M. (1968), "The structure of utility functions", *Review of Economic Studies*, **35**, 367–390.

HAMMOND, P. J. (1981), "Ex-ante and ex-post welfare optimality under uncertainty", *Economica*, **48**, 235–250.

HAMMOND, P. J. (1983), "Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty", in Pattanaik and Salles (1983) 175–206.

HAMMOND, P. J. (1987), "On reconciling Arrow's theory of social choice with Harsanyi's fundamental utilitarianism", in Feiwel (1987).

HARSANYI, J. (1953), "Cardinal utility in welfare economics and in the theory of risk-taking", *Journal of Political Economy*, **61**, 434–435. (Reprinted in Harsanyi (1976), 3–5.)

HARSANYI, J. (1955), "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility", *Journal of Political Economy*, **63**, 309–321. (Reprinted in Harsanyi (1976), 6–23.)

HARSANYI, J. (1976) *Essays on Ethics, Social Behavior, and Scientific Explanation* (Dordrecht: Reidel).

HARSANYI, J. (1977), "Morality and the theory of rational behaviour", *Social Research*, **44**. (Reprinted in Sen and Williams (1982), 39–62.)

HOOKER, C., LEACH and McCLENNEN, E. F. (eds) (1978) *Foundations and Applications of Decision Theory, Volume 1* (Dordrecht: Reidel).

HORGAN, T. (1981), "Counterfactuals and Newcomb's problem", *Journal of Philosophy*, **78**, 331–356.

JEFFREY, R. C. (1971), "On interpersonal utility theory", *Journal of Philosophy*, **68**, 647–656.

JEFFREY, R. C. (1983) *The Logic of Decision* (Second Edition) (Chicago: University of Chicago Press).

JEFFREY, R. C. (1987), "How to probabilize a Newcomb problem", in Fenton (1988), 241–251.

JONES-LEE, M. W. (ed) (1982) *The Value of Life and Safety* (Amsterdam: North-Holland).

LAFFONT, J. J. (ed) (1979) *Aggregation and Revelation of Preferences* (Amsterdam: North-Holland).

LEWIS, D. (1979), "Prisoners' dilemma is a Newcomb Problem", *Philosophy and Public Affairs*, **8**, 235–240.

LEWIS, D. (1981), "Causal decision theory", *Australasian Journal of Philosophy*, **59**, 5–30.

NOZICK, R. (1969), "Newcomb's problem and two principles of choice", in Rescher (1969) 114–146.

PATTANAIK, P. K. and SALLES, M. (1983) *Social Choice and Welfare* (Amsterdam: North-Holland).

RESCHER, N. (ed) (1969) *Essays in Honor of Carl G Hempel* (Dordrecht: Reidel).

SAVAGE, L. J. (1972) *The Foundations of Statistics* (Second Edition) (New York: Dover).

SEN, A. K. (1976), "Welfare inequalities and Rawlsian axiomatics", *Theory and Decision*, **7**, 243–262. (Reprinted in Butts and Hintikka (1977).)
SEN, A. K. and WILLIAMS, B. (eds) (1982) *Utilitarianism and Beyond* (Cambridge: Cambridge University Press).
SIKORSKI, R. (1960) *Boolean Algebras* (Berlin: Springer-Verlag).
SKYRMS, B. (1982), "Causal decision theory", *Journal of Philosophy*, **79**, 695–711.