CLASSIC AND CURRENT NOTIONS OF " MEASURABLE UTILITY "

I

It is ten years since von Neumann and Morgenstern, in their famous aside to the economic profession, announced they had succeeded in synthesising " measurable utility." That feat split their audience along old party lines. It appeared that a mathematician had performed some elegant sleight-of-hand and produced, instead of a rabbit, a dead horse.

The most common reaction was dismay. To " literary " economists who had freshly amputated their intuitive feelings of cardinal utility at the bidding of some *other* mathematicians, it seemed wanton of von Neumann and Morgenstern so soon to sprinkle salt in their wounds with the statement : [1] " It can be shown that under the conditions on which the indifference curve analysis is based very little extra effort is needed to reach a numerical utility." To others, who had said all along that surgery was unnecessary, the verdict was no surprise but still welcome, coming as it did from an unexpected (non-Cambridge) source. But before long both these groups had joined in expressing doubts that von Neumann and Morgenstern had succeeded in doing what (these readers believed) they had set out to do. The spokesman for the " cardinalists," interpreting their cause as his own, was forced to conclude that they " seem to me to have done as much harm as good to the cause to which they have lent their distinguished aid." [2]

However, it is now clear that the impression that von Neumann and Morgenstern were leading a reactionary movement was erroneous. Their cause, if it can be so dignified, is a new one, not that to which Professor Robertson alluded. 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.

In the latest writings the theory has been formulated un-

[1] Von Neumann and Morgenstern, *The Theory of Games* (Princeton, 1944), p. 17.

[2] Professor D. H. Robertson, *Utility and all That* (London, 1952), p. 28.

ambiguously, so that the subject presents little difficulty to one approaching it now for the first time. This article is directed, instead, at readers who came early to the controversy, and who followed the theory in its various stages closely enough to become thoroughly confused.

By concentrating their discussion on the general concept of "measurability," von Neumann and Morgenstern unfortunately obscured the unique features of their particular construction. Later expositions have tended to follow them in this,[1] or to stress the empirical content of the von Neumann–Morgenstern results.[2] Very little attention has been given to the major source of misunderstandings: ambiguity concerning the differences in derivation and application between the new notion of "measurable utility" and the concept of the same name implied in the writings, say, of Marshall and Jevons. This article will attempt to distinguish clearly between the two concepts, chiefly by examining the different operations by which they are defined and tested.[3]

This procedure may provide some valuable exercise in the use of the operational approach, which in economic literature has been honored chiefly in footnotes. This approach regards the basic definition of a technical concept in scientific usage as: "What is measured by" a particular set of operations. Two different sets of operations are presumed to measure two different "things," although under certain conditions (discussed later) it is justifiable to treat the two concepts as identical. A scientific proposition is operationally meaningful if definite conceivable results of given operations are defined which would *refute* the statement; if it does not *restrict* the class of results which are to be expected, it cannot be useful for scientific purposes. The meaningfulness of concepts and propositions is a necessary,

[1] This is the only shortcoming of the otherwise excellent article by Alchian, "The Meaning of Utility Measurement," *American Economic Review*, March 1953, p. 26. The present paper may serve as a complement to Alchian's.

[2] Friedman and Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, August 1948, p. 279. Mosteller and Nogee, "An Experimental Measurement of Utility," *Journal of Political Economy*, October 1951, p. 371. I have also benefited from reading as yet unpublished papers by Professors Bishop, Marschak and Allais.

[3] This paper was originally written as the first chapter in a thesis, entitled "Theories of Rational Choice Under Uncertainty: The Contributions of von Neumann and Morgenstern," submitted for undergraduate honors at Harvard University, April 1952. The thesis was written under the valuable guidance of Professor John Chipman. I am also greatly indebted to Professors Paul Samuelson, Robert Bishop, Oskar Morgenstern and Frederick Mosteller for the opportunity to discuss problems and to read unpublished writings on the subject, and to Mr. Nicholas Kaldor for his comments.

though not a sufficient, condition for their scientific usefulness.[1] This point of view will prove useful in the concluding section in clarifying the *difference* in meaning of two concepts bearing the same name.

The next section will describe the similarities between the old and new approaches to a " cardinal utility " and will present, in advance, some of the conclusions to be drawn as to their points of contrast. The next two parts examine the operational bases of the two theories, and the final section will analyse in detail the peculiarities of the von Neumann–Morgenstern construction.

II

Suppose that a man who prefers A to B, B to C and A to C must choose between having B for certain or having a " lottery ticket " offering A with probability p or C with probability $1 - p$.[2] Without asking him outright, is it possible to predict his choice? If so, what sort of data are necessary?

Economists of the school of Jevons, Menger, Walras and Marshall, on the one hand, and on the other those following von Neumann and Morgenstern would answer " Yes " to the first question. But their predictions would be based on quite different types of data.

For von Neumann and Morgenstern, it would be necessary to observe the man's behavior in other risk-situations, involving different outcomes or the same outcomes with different probabilities. The older economists, of whom we will take Marshall as typical, would ask no knowledge of his other risk-behavior. They assumed it possible, by observation or interrogation, to discover a man's intensities of liking for sure outcomes; on the basis of this knowledge alone, they were ready either to predict or to prescribe his choice between prospects.

This divergence is concealed by the fact that both schools would summarise the results of their investigations in the same symbolic shorthand, arriving at expressions that are formally

[1] These concepts and the general point of view were first formulated explicitly by Percy W. Bridgman, in *The Logic of Modern Physics* (New York, 1927), who declared them to be implicit in the thinking of modern physicists. The terms, and the emphasis on restrictiveness and refutability of propositions, have become familiar to economists largely through Samuelson's *Foundations of Economic Analysis* (Cambridge, 1948), but the other main propositions are less well known.

[2] A " lottery ticket " of this sort, offering a set of alternative outcomes with stated probabilities summing to unity, will hereafter be known as a *prospect*. If one outcome is offered with unit probability, *i.e.*, with no uncertainty, it will be known as a *sure outcome*.

identical. Under both procedures the results of experiment would be expressed by assigning a triplet of numbers, U_a, U_b and U_c, to the three outcomes, with the property: $U_a > U_b > U_c$. This triplet is a utility index for the three outcomes, since their order of magnitude reflects the order of preference. Next, both Marshall and von Neumann–Morgenstern would form the expression:

$$(1) \qquad 1 \, . \, U_b \gtreqless p \, . \, U_a + (1 - p)U_c$$

where p and $1 - p$ are the respective probabilities of A and C. Each side of this relationship is a sum of the utility numbers corresponding to the outcomes of a given prospect multiplied by their respective probability numbers. Since the probabilities sum to unity, the result is a weighted arithmetic mean of the utilities, variously known as the mathematical expectation of utility, the expected utility, the moral expectation, moral expectancy, actuarial value of utility and the first moment of the utility–probability distribution.[1]

In each case the prediction (or advice) would have the man choose the prospect with the highest mathematical expectation of utility. Or, if the two sides of the relationship (1) were equal, he should be indifferent between the two prospects. A man whose behavior conformed to this rule could be said to be "maximising the mathematical expectation of utility."

With this much similarity between the two approaches, it is natural that they should commonly be confused. Yet the most misleading point of similarity remains: the fact that in both cases "utility" is said to be "measurable." The necessity of this assumption is seen more clearly if the left-hand side of expression (1) is rewritten $p \, . \, U_b + (1 - p)U_b$ and terms collected to form the relationship:

$$(2) \qquad p(U_a - U_b) - (1 - p)\,(U_b - U_c) \gtreqless 0$$

This is merely relationship (1) in a different form. The rule would now have the man accept the prospective offering A or C if the left-hand side of (2) were positive, reject it if the left-hand side were negative, or be indifferent between it and the sure outcome B if the left-hand side were equal to zero. The important point here is that relationship (2) shows clearly that the rules rely on comparing *differences* in utility.

[1] Of these, "mathematical expectation of utility" and its shorter form, "moral expectation," will be used below. Both must be carefully distinguished from the "mathematical expectation of *money*," which is a weighted sum of the money outcomes, rather than of their utility numbers.

If only preferences were known, the triplet of utility numbers could be replaced by another with the same ordinal relationships. In general, such a monotonic transformation would not preserve equality, or given inequalities, among differences between utility numbers. The rule of maximising expected utility would lead to prediction or advice which would depend on the particular index used, and if preferences were the only guide, the choice of index would be arbitrary. The rule would be meaningless, therefore useless.

In order for the rule to give definite results, it would be necessary to find some "natural" [1] operation that would give meaning to differences in utility numbers, hence to the numerical operations implied by the rule. The new index, summarising the results of the additional operation as well as preferences, would belong to a more restricted set of indices than the set of all ordinal utility indices. Any two indices in which corresponding differences as well as absolute utilities satisfied the same inequalities would be related by a linear, and not merely any monotonic, transformation.[2] It is, then, necessary to find some aspect of behavior that can be described only by a set of numbers determined up to a linear transformation: a set, moreover, which is one of those expressing preferences. So much is necessary in order for the rule of maximising expected utility to be meaningful. Its usefulness, if any, must depend on the particular aspect of behavior which serves this purpose, if one can be found.

To say that both the Marshallian and the von Neumann–Morgenstern theories require a "measurable utility" is precisely to say that they require a utility index determined up to a linear transformation. At this point the similarity ends. In general, the order of magnitude of the differences between corresponding numbers would be different for the two indices; therefore, predictions based on the rule of maximising moral expectation

[1] Von Neumann and Morgenstern use this term to signify an operation other than numerical or logical manipulation of a mathematical model. A mathematical model is useful if the results of a "natural" operation can be correlated with numbers in such a way that numerical operations can symbolise and substitute for the "natural" operation.

[2] For an excellent exposition of the concepts of linear and monotonic transformations, the reader is referred to the article by Alchian, *op. cit.* Briefly, two indices are related by a linear transformation if for every point x on one index, the corresponding point y on the other index satisfies a relationship of the form: $y = ax + b$, where a and b are constants. The two indices differ only with respect to scale and origin.

If the difference between two numbers in one index is greater than, less than or equal to the difference between two other numbers, the corresponding differences in the other index will have the same ordinal relationship.

would differ for the two approaches. Moreover, it might be possible to find a measurable index by one method and not the other. Even if both should " exist," they would be in general monotonic, and not linear, transformations of each other.

III

Such theorists as Jevons, Menger, Walras and Marshall conceived of the crucial natural operation in the measurement of utility as taking place within the mind of a subject; it was a process of weighing introspectively the amounts of " satisfaction " associated with different outcomes.

Such an operation appeared more of an objective basis for theory to them than it would to modern economists. In their view, in the realm of reasonable men one man's introspection was as good as another's, and the theorist's own internal calculations were likely to correspond roughly to those of his subject; to this extent the results of the subject's operation were " observable." However, if challenged to produce less-subjective evidence, it would undoubtedly have occurred to Jevons that the most natural way to obtain the results of the man's introspective measurements would be to ask for them.

The first rough outline of the subject's pattern of " satisfaction " would emerge from an " indifference-map " experiment, in which he is asked to rank the events, A, B and C in order of preference. If he can compare the events and if his preferences are transitive (" consistent "), *e.g.*, if he prefers A to B, B to C and A to C, the results of this experiment are summarised by any triplet of numbers satisfying: $U_a > U_b > U_c$. This triplet is a " non-measurable " utility index, determined up to a monotonic transformation.

In what we will call a " Jevonsian "[1] experiment the man would next be asked to rank his preferences of A to B and his preference of B to C. If he finds that he can state, for example, that his preference of A to B exceeds his preference of B to C, we could summarise this information by any triplet of numbers satisfying the two inequalities: (*a*) $U_a > U_b > U_c$, and (*b*) $U_a - U_b > U_b - U_c$.

Finally, if A and B were sums of money, we could ask the man to vary the sum of money represented by B until he could tell us

[1] This name is suggested by J. C. Weldon, who points out that it implies no more than that Jevons assumed that preferences could be directly compared. " A Note on Measures of Utility," *Canadian Journal of Economics and Political Science*, May 1950, p. 230.

that he found his preference of A to B' equal to his preference of B' to C. If he finds such a B', then the results of this last operation would be expressed by any triplet of numbers satisfying the relationships : (*a*) $U_a > U_b > U_c$, and (*b*) $U_a - U_b = U_b - U_c$. Any two triplets obeying these relationships must be related by a linear transformation; they represent utility indices differing only by scale and origin.

The Jevonsian index for the individual, if one can be found, is thus " measurable "; which in this case means nothing more or less than that the subject was able to give consistent answers to these particular questions. It might be objected that in fact subjects will be unable to answer the questions, or will answer them inconsistently. This is an empirical matter. If the events were the possession of (*a*) one million dollars, (*b*) two dollars, and (*c*) one dollar, it seems likely that most people would answer the question (and, moreover, would state specifically that their preference of A to B exceeded their preference of B to C).[1] For such people, the notion of a cardinal utility index would not be " meaningless "; if it had no other meaning, it might at least imply that their answers to this sort of question could be predicted. Inconsistency is to be expected, particularly with respect to utility differences that are almost equal. But inconsistency also appears (in lesser degree) in the " indifference-map " experiment; in each case the most important information gained concerns choices which the subject finds easy to make.[2]

The more damaging attack has been on the usefulness of the method, though here again the case is not conclusive. If the only " consistency " discovered were consistency of answers with other answers, the results would be trivial. But Marshall and his predecessors regarded such answers as revealing the subject's internal measurements of satisfaction.[3] Since they believed the man based his decisions to act on the results of this introspective operation, they hoped to use the results of a Jevonsian experiment to predict his decisions.

As the ordinalists have demonstrated, decisions in the market-place under conditions of certainty can be predicted on the basis

[1] At least, they would probably do so if asked point-blank and not given time for doubts as to whether the question " meant " anything (induced, perhaps, by the writings of Samuelson). This could be made part of the conditions of the experiment.

[2] The above discussion follows Weldon, *op. cit.*

[3] Such information might by itself be of interest in welfare economics; it might, though it need not, influence the evaluations on which a social-welfare function must be based.

of the " indifference-map " experiment alone. But this is not true of behavior under uncertainty or risk. With a Jevonsian utility index, on the other hand, it is possible to frame meaningful hypotheses placing definite restrictions on observable behavior in risk-situations.

The particular rule which Marshall and Jevons proposed (rather more for normative purposes than descriptive) was that the " rational " man would maximise the mathematical expectation of utility. In terms of the expression (1) cited earlier :

$$(1) \qquad U_b \gtreqless p \,.\, U_a + (1 - p)U_c$$

the rational man should (would) choose the prospect if the right-hand side were greater, the sure outcome B if the left-hand side were greater. If U_b, the utility that can be had for certain if the prospect is rejected, is regarded as the opportunity cost of the prospect, then the expression (2) :

$$(2) \qquad p(U_a - U_b) - (1 - p)\,(U_b - U_c) \gtreqless 0$$

represents the mathematical expectation of *gain* (measured in utility) associated with the prospect. The first term is the amount of utility that the man stands to win by accepting the gamble (in excess of the utility cost of the gamble) multiplied by the probability of winning, and the second term is the amount of utility he stands to lose multiplied by the probability of losing. The man should take any gambles whose expectation of gain is positive, reject all whose expectation of gain is negative, be indifferent to those whose expectation of gain is zero.

One point about this procedure must be emphasised, for it is in sharp contrast to that of von Neumann and Morgenstern. The utility index, and its measurability, on which the Marshallian predictions were based was not derived from any risk-behavior, and did not depend on any sort of consistency in that behavior. The rule of maximising expected utility on the basis of a Jevonsian index led to prediction, or prescription, of a man's choices among prospects without any previous observation of his behavior in the face of risk.

Actually, in the main field of consumer behavior characterised by risk—gambling—Marshall was not sanguine about the usefulness of the rule as a descriptive hypothesis. He took it as a universal empirical law that answers in the Jevonsian experiment would reveal diminishing marginal utility. In other words, if A, B and C are three sums of money such that $A > B > C$, and if $A - B = B - C$, then he assumed that corresponding utility

numbers would satisfy the inequality: $U_a - U_b < U_b - U_c$. A "fair" gamble is defined as one in which the mathematical expectation of *money* gain is zero, expressed by:

$$(3) \qquad p(A - B) - (1 - p)(B - C) = 0$$

where p is the probability of winning A, $1 - p$ the probability of winning C and B is the cost of the gamble (in the above case, p must equal $\frac{1}{2}$). But, granted decreasing marginal utility, the corresponding expectation of *utility* gain is negative, so the rational man would never accept a fair gamble, or, *a fortiori*, an unfair gamble.

As Marshall was well aware, people did accept fair and even unfair gambles; but this behavior disputed their rationality, not (the curvature of) their utility index. The latter was established once and for all by tests that did not involve risk. Because of the existence of "pleasures of gambling" [1] (which Marshall measured by the acceptance of unfair bets), Marshall would have rejected the observation of risk-behavior as an alternative operation for measuring people's intensities of liking for outcomes.

The particular Marshallian rule governing risk-behavior is not implied by his concept of utility or by the methods of measuring it. An early form of the rule is stated by Jevons: [2]

> "If the probability is only one in ten that I shall have a certain day of pleasure, I ought to anticipate the pleasure with one-tenth of the force which would belong to it if certain. In selecting a course of action which depends on uncertain events, as, in fact, does everything in life, I should multiply the quantity of feeling attaching to every future event by the fraction denoting its probability."

The reliance of this approach on measurability (of "quantity of feeling") is obvious; but it is equally obvious that the measurability of "pleasure," and even the general principle that likings for prospects should be based on likings for outcomes and their probabilities, does not imply this particular rule of decision-making. On the basis of a given Jevonsian index, Marshall or Jevons could just have easily proposed that the rational man base

[1] Unfair gambling could be "rationalised" if introspective tests revealed that the happiness derived from gambling outweighed the "expected loss of satisfaction" implied by the odds. However, Marshall wished to retain the normative connotation of "rationality" at the expense of predictive value. In his view, the pleasures of gambling "are likely to engender a restless, feverish character, unsuited for steady work. . . ." (*Principles of Economics* (London, 1925), Mathematical Appendix, Note IX, p. 843.) Granted that marginal utility was decreasing, and that pleasures of gambling could be ignored because "impure," then unfair gambling was unequivocally irrational, an "economic blunder."

[2] W. Stanley Jevons, *The Theory of Political Economy* (London, 1911), p. 36.

his preference on the mode, the median, the range, variance or other properties of the distribution of utilities. These rules would have been just as meaningful, and possibly more useful (especially if they took into account measures of " risk " as well as " central tendency "). In fact, it was the feeling that the emphasis on mathematical expectation was arbitrary and unrealistic which led to the decline of the concept even before doubts arose that a measurable utility could be discovered to make it meaningful.

IV

What von Neumann and Morgenstern asserted, in their famous digression,[1] was the possibility that the notion of maximising the mathematical expectation of utility might (*a*) be made meaningful, and (*b*) describe a wider range of risk-behavior than in its old usage, *if " utility " were measured (defined) in a special way.* Since they were concerned only with risk-behavior, the operation they proposed was the observation of choices in risk-situations. If a person's preferences among *prospects*—described merely in ordinal terms—should satisfy certain, apparently weak, axiomatic restrictions, then von Neumann has proved that it would be possible to find a set of numbers which could express these preferences in a particularly convenient way.

This set of numbers would be *a* utility index, because it would be one among all the sets of numbers (related by monotonic transformations) expressing the person's preferences (*not* " intensities of preference " or " quantities of feeling ") among sure outcomes. The novelty would be that this same set of numbers, applying explicitly only to sure outcomes, could also summarise the person's preferences among prospects. In a complete description of the individual's entire preference-structure, it would be unnecessary to list prospects separately or to record explicitly his preferences among prospects; these preferences would be known, through observation, but they could be expressed implicitly by the numbers attached to sure outcomes.

Clearly, the class of indices which could express with such economy preferences both among sure outcomes and among prospects must be smaller than the class of all ordinal utility indices (in most of which it would be necessary to list prospects individually). In fact, it turns out that all indices with this property,

[1] The theory which follows occupies only a few pages in the introduction to their book, and plays no role in the theory of games. The latter theory requires a commodity which is not only measurable but intercomparable and freely transferable, so pay-offs are expressed in money, not in von Neumann–Morgenstern " utility."

if any exist, will be related by linear transformations. Yet the index is not " measurable " in the sense that it is correlated with any significant economic quantity such as quantity of feeling or satisfaction, or intensity, such as intensity of liking or preference. It is derived from *choices*, and describes only *preferences*. It would be " cardinal " (" measurable ") only to the extent that the numerical operation of forming mathematical expectations on the basis of these numbers would be related to observable behavior, so as to be empirically meaningful.

Von Neumann and Morgenstern might simply have proposed the empirical hypothesis that an index of the desired sort could be found for certain individuals. However, this proposition, which we will call the Hypothesis on Moral Expectations, has little inherent plausibility. The major feat of von Neumann and Morgenstern is to show that the Hypothesis on Moral Expectations is *logically equivalent* to the hypothesis that the behavior of given individuals satisfies certain axiomatic restrictions. Since these axioms appear, at first glance, highly " reasonable," the second hypothesis seems far more intuitively appealing than the equivalent Hypothesis on Moral Expectations. It is thus more likely to be accepted on the basis of casual observation and introspection, although the two hypotheses would both be contradicted by exactly the same observations.

Most expositions follow von Neumann and Morgenstern in focusing all attention on the second hypothesis, *i.e.*, on the empirical relevance of the axioms. Once this is accepted, the Hypothesis on Moral Expectations " goes along free " in the form of the Theorem on Moral Expectations, which states conditionally that *if* an individual's behavior conforms to the axioms, a von Neumann–Morgenstern index can be computed for him (this proposition rests on logic rather than observation, and it has been established by several different proofs). Empirical test of the proposition is thus displaced to the axioms which imply it. The logical relationship of the axioms to the Hypothesis is usually left obscure, for the demonstration is too difficult for most readers.[1]

[1] Von Neumann and Morgenstern did not present a proof deriving the Theorem from the axioms until the second edition of *The Theory of Games* (1947); they describe it, with terrific understatement, as " rather lengthy and may be somewhat tiring for the mathematically untrained reader " (p. 617). A different, slightly easier proof, is given by Marschak in " Rational Behaviour, Uncertain Prospects, and Measurable Utility," *Econometrica*, April 1950. A genuinely simple proof has finally been presented by Samuelson, in " Utility, Preference, and Probability," abstract of paper given before the conference on *Les Fondements et Applications de la Theorie du Risque en Econometrie*, 1952.

Therefore the reader must generally take it on faith that behavior violating a particular axiom conflicts with the possibility of finding a von Neumann–Morgenstern index. Instead, we will follow the straighter, though less persuasive, route of describing how the Hypothesis on Moral Expectations would be tested directly.

We can state the Hypothesis in the following form. For a given individual (it is asserted that) a set of numbers *exists* (*i.e.*, can be found) with the two properties : (1) it is one of the sets expressing the individual's actual preferences among sure outcomes (*i.e.*, it is one of his ordinal utility indices); (2) numbers are assigned to sure outcomes in such a way that, if " moral expectations " of *prospects* were computed on the basis of these numbers, one prospect would have a higher moral expectation than another if, and only if, the person actually preferred the former to the latter, and two prospects would have the same moral expectation if and only if the person were indifferent between them.

If, from the set of all utility indices (related by monotonic transformations) one index can be found such that " moral expectations " computed on the basis of this particular set of " utilities " arrange prospects according to an individual's actual preferences among them, then any other index related to the first by a linear (not merely by any monotonic) transformation will also have this property. Thus, if one such index exists, an infinite set will exist : though still a tiny subset of all indices expressing ordinal preferences among outcomes. The Theorem on Moral Expectations states that such a set does exist, if and when the axioms apply. The Hypothesis states that the index actually does exist for given persons.

Having found such an index, we could submit it to a monotonic increasing transformation—*e.g.*, we could take the square or the log of each number—and the resulting set of numbers would be a perfectly valid utility index of outcomes. But it would not serve any more as a utility index of prospects as well; it would no longer be true that moral expectancies would correspond to the individual's actual preferences among prospects.

Our approach will consist of trying to find an index (hypothetically) with the two properties specified, noting in the process the type of behavior which would make this impossible. The " operational content " of the theory should be most obvious from this point of view, since it is intimately related to the body of behavior " ruled out " by the Hypothesis. The greater the amount and

importance of this behavior, the more powerful does the Hypothesis appear, though the less immediately plausible.

The basic operation in deriving a von Neumann–Morgenstern utility index is the observation of an individual's behavior in the very simplest situation involving risk : a choice between a sure outcome and a prospect involving two possible outcomes with given probabilities. The essential restriction the Hypothesis puts on behavior is that, by observing a person's choices in situations of this simple type, it must be possible to predict his choices among sets of prospects each offering a multitude of prizes with complex odds (some of the prizes possibly being other prospects). In the discussion below, the notation $(A, p;\ B)$ signifies a prospect offering outcome A with probability p or B with probability $1 - p$.

To fix the origin and unit of the utility index we seek, we assign arbitrary numbers to two outcomes (order of magnitude in order of preference); this guarantees that the index, if we can find one, will be unique. For example, let us assign to the money-sums \$1,000 and \$0 the utility numbers 10 and 0 : *i.e.*, $U_{1000} = 10$, $U_0 = 0$. Now we consider a third sum, say, \$500, which the individual ranks between the first two; the problem is to find a utility number U_{500} that satisfies the Hypothesis on Moral Expectations, consistent with his preferences and with the two numbers already assigned arbitrarily. The crucial datum in the procedure is the probability $\dot{p}$ at which the person is indifferent between having \$500 with certainty or a prospect (\$1,000, $\dot{p}$; \$0).[1] Suppose that this $\dot{p}$ is $\frac{8}{10}$; *i.e.*, he tells us, or we observe, that he is indifferent between \$500 and (\$1,000, $\frac{8}{10}$; \$0). The Hypothesis on Moral Expectations then implies that it is possible to find a number U_{500} with the two properties :

(4.1) $0 < U_{500} < 10$ (since he prefers \$1,000 to \$500 and \$500 to \$0)

(4.2) $$1 \,.\, U_{500} = \tfrac{8}{10} \,.\, 10 + \tfrac{2}{10} \,.\, 0$$

Obviously, such a number *can* be found : $U_{500} = 8$. So the Hypothesis has passed the first test.

Even in this first application, the Hypothesis was not tautologous. It was conceivable that the individual would prefer the certainty of \$500 to any prospect (\$1,000, p ; \$0) for *any* p what-

[1] The axioms require that he be indifferent at one and only one p. Tests have already shown that this perfect consistency is never encountered, but " indifference " might be defined stochastically (*e.g.*, if an individual rejected a prospect with given odds as often as he accepted it, he might be said to be " indifferent " to it).

ever : perhaps from extreme conservative principles or moral scruples against gambling. The Hypothesis would then imply that it was possible to find a number U_{500} satisfying both the following two relationships :

$$(5.1) \qquad 0 < U_{500} < 10$$

and

$$(5.2) \quad U_{500} > p \,.\, 10 + (1 - p) \,.\, 0, \text{ for } all\ p,\ 0 < p < 1.$$

But no such number exists; for any given U_{500} satisfying (5.1), there would exist a p, such that (5.2) would not hold. Therefore the Hypothesis would be contradicted.

Similarly, the Hypothesis would be contradicted if the individual should prefer any prospect (\$1,000, p ; \$0) to the certainty of \$500; say, from an obsession with gambling.

It might seem that such behavior might well occur, thus rejecting the Hypothesis on the basis of one observation. But proponents of the Hypothesis could point out that it is unusual to have such " absolute " likes or dislikes : to feel so strongly either for or against gambling as to ignore entirely the relative stakes and odds.[1] They might suggest that such behavior, though it may exist, is statistically unimportant, so that it is reasonable to hypothesise that there will be *some* $\dot{p}$ at which the subject will be indifferent. A man with a marked taste for security might pick $\dot{p} = 9{,}999/10{,}000$; a born gambler might indicate $\dot{p} = 1/1{,}000$. In either case it would be possible to find a number U_{500} consistent with these preferences.

Thus, the Hypothesis puts very weak limitations in this initial application to the man's preferences among risky alternatives. The drastic test is to investigate whether or not his *other* choices will be " consistent " with this first choice. Let us return to our original result, $U_{500} = 8$. On the basis of our single observation (fixing $\dot{p}$ at $\frac{8}{10}$) we must be able to predict the individual's choice among any set of prospects involving the three outcomes, \$1,000, \$500 or \$0, with any probabilities. Given any set of prospects, we simply compute the moral expectations of each on the basis of the utility numbers (two of which, in this case, were fixed arbitrarily and the third derived from a single observed choice), and pick the prospect with the highest moral expectancy. No rationale for this procedure has been given here. It is not suggested that the individual makes his choice by a similar calculation. We are merely examining the implications

[1] As John Chipman has put it, this is the " every man has his price " axiom.

of the hypothesis that it is possible to describe his behavior "as though" he did.

Thus, we compute the moral expectation of the prospect ($1,000, $\frac{1}{2}$; $0) as : $\frac{1}{2} . 10 + \frac{1}{2} . 0 = 5$. If, when confronted with the choice between this prospect and the certainty of $500 ($U_{500} = 8$), our subject does not definitely prefer the latter, then it is not true that moral expectations computed on the basis of our utility numbers arrange prospects according to the individual's actual preference; our triplet does not have properties of a von Neumann–Morgenstern index. More than that, if this triplet is not one of those whose existence is implied by the Moral Expectations Hypothesis, *then no such triplet can be found*, and the hypothesis is thereby invalidated. For, once two numbers had been arbitrarily chosen, the third one was uniquely determined by our initial observation; [1] any other value for U_{500} would be inconsistent (in terms of our hypothesis) with that particular choice.

If, on the contrary, no serious [2] inconsistency appears, we can proceed to find utility numbers for other sums of money. If we observed that the subject was indifferent between $200 and ($500, $\frac{1}{4}$; $0), we would define $U_{200} = 2$. Our set of utility numbers corresponding to $0, $200, $500 and $1,000 is now 0, 2, 8, 10. If these are the unique set implied by the Hypothesis, then the individual should be indifferent between a 50–50 chance of $0 or $1,000, and a 50–50 chance of $200 or $500, since : $\frac{1}{2} . 0 + \frac{1}{2} . 10 = \frac{1}{2} . 2 + \frac{1}{2} . 8$. If, in fact, he prefers one to another, then the existence theorem is contradicted; the axioms on which it may be based do not apply to this individual.

More complicated tests can be devised. One of the "prizes" in a prospect might be another prospect, say, a "lottery ticket" offering a $\frac{4}{5}$ chance of $1,000 and a $\frac{1}{5}$ chance of $0; if the other prize is $200, the two prizes being offered at equal odds, this would appear in our notation : ($200, $\frac{1}{2}$; ($1,000, $\frac{4}{5}$; $0)). This "complex" prospect might be compared to the "simple" prospect : ($500, $\frac{5}{8}$; $0). The person "should" be indifferent between them, since :

$$\frac{1}{2} . 2 + \frac{1}{2}(\frac{4}{5} . 10 + \frac{1}{5} . 0) = \frac{5}{8} . 8 + \frac{3}{8} . 0$$

A new test would be to confront the subject with a choice between the above complex prospect and the simple prospect with

[1] Since indices satisfying the Hypothesis are determined "up to two arbitrary constants," the specification of two values determines the index uniquely.

[2] In a real experiment we would have to decide in statistical terms what to regard as a "reasonable" approximation to consistency with the Theorem.

three prizes : (\$1,000, $\frac{2}{5}$; \$200, $\frac{1}{2}$; \$0, $\frac{1}{10}$). Suppose that U_{200} is yet to be computed, and that the individual is found to prefer the complex "lottery ticket" to the above simple one. Then the Hypothesis would imply that a number U_{200} can be found satisfying both :

(6.1) $0 < U_{200} < 10$, and

(6.2) $\frac{2}{5} \cdot 10 + \frac{1}{10} \cdot 0 + \frac{1}{2} \cdot U_{200} < \frac{1}{2}(\frac{4}{5} \cdot 10 + \frac{1}{5} \cdot 0) + \frac{1}{2} \cdot U_{200}$

Since (6.2) implies $U_{200} < U_{200}$, it is clearly impossible to find a number with the desired properties.

Von Neumann and Morgenstern's controversial Axiom 3 : C : b[1] rules out this type of behavior by assuming that a person will be indifferent between two prospects which are derivable from each other according to the rules of probabilities. By application of these rules, any complex prospect offering other prospects as prizes may be reduced to a simple prospect, and the axioms requires the individual to be indifferent between this derived prospect and the original one. This implies that the individual is indifferent to the number of steps taken to determine the outcome. On the contrary, a sensible person might easily prefer a lottery which held several intermediate drawings to determine who was still "in" for the final drawing; in other words, he might be willing to pay for the possibility of winning intermediate drawings and "staying in," even though the chances of winning the pot were not improved thereby. A longer time-period of suspense would usually also be involved, but it need not be. The crucial factor is "pleasure of winning," which may be aroused by intermediate wins even if one subsequently fails to receive the prize. Many, perhaps most, slot-machine players know the odds are very unfavorable, and are not really motivated by hopes of winning the jackpot. They feel that they have had their money's worth if it takes them a long while to lose a modest sum, meanwhile enjoying a number of intermediate wins—which go back into the machine to pay for the pleasure of the next win. Von Neumann and Morgenstern single out axiom 3 : C : b, which excludes this type of behavior, as the "really critical" axiom [2]—"that one which gets closest to excluding a 'utility of gambling.'" [3]

The final major test of the Hypothesis would be to give the subject a choice between two such prospects as (\$500, p; \$1,000) and (\$200, p; \$1,000), where p is the same in each and where \$500

[1] Von Neumann and Morgenstern, *op. cit.*, p. 26. [2] *Ibid.*, p. 632. [3] *Ibid.*, p. 28.

is preferred to \$200. For any p he must prefer the first to the second. If, for example, he was indifferent between them at some $p = P$, the Hypothesis would imply that there was a U_{500} such that:

$$(7.1) \quad 2 < U_{500} < 10, \quad \text{and}$$

$$(7.2) \quad P \,.\, U_{500} + (1 - P)10 = P \,.\, 2 + (1 - P)10.$$

Together these imply that $U_{500} > 2$ and $U_{500} = 2$, which can be true of no number (we are assuming in this example that U_{500} has not already been determined by some other experiment).

It is the "Strong Independence Axiom" ruling out this sort of preference which Samuelson has emphasised, presenting it as the "crucial" axiom.[1] It seems rather hard to justify this emphasis, since the axiom seems indubitably the most plausible of the lot. After all, all of the axioms are necessary to the final result, and this particular one is almost impregnable (even people who did not follow it in practice would probably admit, on reflection, that they should) whereas others (such as $3 : C : b$) are contradicted by much everyday experience. One might almost suspect Samuelson, who counts himself a "fellow traveller"[2] of the von Neumann–Morgenstern theory, of using the axiom (his invention) as a man-trap, luring critics past the really vulnerable points to waste their strength on the "Independence Assumption."[3]

In all this it has been emphasised that the Hypothesis on Moral Expectations sets a double condition for an acceptable index, the first part being that it must be one of the individual's ordinal utility index. Some critics seem to have overlooked this; for example, I. M. D. Little:[4]

> "Suppose that . . . we had given C, A, B, the utility numbers $10\frac{4}{5}$, 10, 9, because the consumer was 'indifferent' between (A certain) and (C with probability $\frac{5}{9}$ or B with probability $\frac{4}{9}$). It follows that . . . if the consumer is given the choice between B and A, A must be taken. In fact B might well be taken."

[1] Samuelson, "Utility, Preference, and Probability," *op. cit.* Also, "Probability, Utility, and the Independence Axiom," *Econometrica*, October 1952, p. 672.

[2] Samuelson, "Probability, Utility, and the Independence Axiom," *op. cit.*, p. 677.

[3] Dr. Alan S. Manne's article, "The Strong Independence Assumption—Gasoline Blends and Probability Mixtures," *Econometrica*, October 1952, p. 665, gives an interesting example of a physical situation in which superposition does not apply, which may be more relevant to linear programming than to the present subject. Although it raises a doubt, I do not think his criticism is really damaging in this context. The argument in the same issue by H. Wold, "Ordinal Preferences or Cardinal Utility?" is definitely invalid.

[4] I. M. D. Little, *A Critique of Welfare Economics* (Oxford, 1950), p. 30.

If, as the last sentence suggests, the consumer preferred B to A (and C to both), we would start the experiment with this information. If we should then observe that he was indifferent between A and $(C, \frac{5}{9};\ B)$, then we would not be able to assign any utility number at all to A, for it would be impossible to find one satisfying the two conditions: $U_a < U_b < U_b$ and $U_c = \frac{4}{9} . B + \frac{5}{9} . C$. This behavior contradicts the Hypothesis, but the conflict would show up in the impossibility of finding a von Neumann–Morgenstern index, not in the index, once having been " certified," turning out to be inconsistent with ordinal preferences. This is a small point, but criticism which may be quite pertinent loses force if framed in a way that suggests the critic has not understood the conditions of the experiment.

Another type of criticism that goes wide of the mark uses examples involving only " utils," with no mention of the sums they represent or the particular observations on which they were based. Baumol, for example, cites two lottery tickets with prizes expressed in utils (*i.e.*, utility units, rather than money).[1] He computes their moral expectations, but asks, " Yet who is to say " that it is " pathological " for the subject to prefer the one with the lower expectation. To this a defender can retort: (*a*) Baumol gives no indication that the utility numbers were correctly derived; (*b*) it would not, of course, be " pathological " in any case; but (*c*) if it happened that the utility numbers were actually derived, for example, from the person's previous choice between the very two prospects cited, then it would be " inconsistency " of a sort usually defined as non-rational for him to switch his choice on this occasion. The crux of the matter is that it is impossible to decide on intuitive grounds whether it is " plausible " to choose the prospect with the higher moral expectation if only utils are cited and if the person's past choices are not known, since an appraisal of " plausibility " must be based on the money sums involved and on the person's pattern of behavior in risk-situations.

We have described above the main types of behavior that conflict with the Hypothesis on Moral Expectations. It is possible to give long lists of factors in risk-situations which would lead to these types of behavior.[2] Among those which have not

[1] William Baumol, " The Neumann–Morgenstern Utility Index—An Ordinalist View," *Journal of Political Economy*, February 1951, p. 65.

[2] Maurice Allais, in " Notes théoriques sur l'Incertitude de l'Avenir et la Risque " (as yet unpublished), and Professor Robert Bishop, in a paper that has not been published as yet, outline these considerations in detail.

been mentioned earlier are: feelings of skill, or, in general, the feeling that the "real" odds are more favorable than the stated odds (*e.g.*, belief in personal luck, or in "winning streaks"); inability to compute compound probabilities, and thus to derive simple prospects from complex ones; influence of the other elements in the risk-situation besides the money prizes and the probabilities—*e.g.*, the atmosphere of the gaming-room. The Theorem could possibly be framed so as to allow for these considerations, but in any practical application they would undoubtedly have some effect.

Whether or not these factors would lead to *serious* inconsistency is open to question; it seems very likely that they would in the field of gambling, but Samuelson suggests that they may be less important in business and statistical problems.[1]

The only laboratory test of the Hypothesis has been performed by Professor Frederick Mosteller, who derived utility curves for a group of subjects on the basis of their choice among simple gambles, and used these data to predict their choices among other and more complicated gambles.[2] The experiment side-stepped pitfalls which could not be avoided in practical application by abstracting from the major sources of inconsistent behavior: (*a*) all probabilities were known; (*b*) all calculations were performed for the subjects and their misconceptions eliminated;[3] (*c*) only small sums were used; (*d*) no social influences or any "other factors" were present; (*e*) behavior was observed only in one special risk-context (and that an artificial one). The fact that Mosteller found only mild consistency despite these "ideal" conditions might be interpreted as distinctly unfavorable to the hypothesis (though, considered in themselves, the results were inconclusive).

[1] Samuelson, "Probability, Utility, and the Independence Axiom," *op. cit.*, p. 677.

[2] Frederick Mosteller and Philip Nogee, "An Experimental Measurement of Utility," *Journal of Political Economy*, October 1951, p. 399.

[3] In the first two sessions subjects were not instructed on computing odds, and calculations were not performed for them. Behavior in these sessions was quite different from behavior in the rest of the experiment. Although these interesting results were not discussed in the article cited, Mosteller informed me that a definite finding of the experiment was that all the subjects behaved very differently before and after they had received lectures on dealing with probabilities. Moreover, their behavior showed a trend factor throughout the experiments as they grew increasingly familiar with the various gambles.

Even if the final conclusions had been much more favorable than they were, these observations would have dictated great caution in extrapolating them to situations outside the laboratory.

V

In deriving a " Jevonsian " utility index, we would begin as in the preceding section by assigning two arbitrary values; since, like the von Neumann–Morgenstern index, it is determined up to a linear transformation (if it can be found at all). As before, we might assign the utility numbers 10 and 0 to the outcomes $1,000 and $0. To find U_{500}, instead of confronting the individual with a choice between prospects, we would ask him to rank his preference of $1,000 to $500 and his preference of $500 to $0. Suppose he should tell us that the two preferences were equal; we would then assign the utility number $U_{500} = 5$. But on the basis of the von Neumann–Morgenstern experiment (let us assume that the same individual was the subject) we assigned the number $U_{500} = 8$. Is there not a conflict here ?

To anyone who has skimmed the literature in this field it will not be obvious that the two sets of results are independent, hence do not conflict, for certain passages, particularly in *The Theory of Games*, gives quite the opposite impression. A close examination of the texts can, in fact, settle the question definitely. Instead of referring immediately to the literature, however, it is rewarding to examine a more general type of analysis, which might have made the issues intelligible to economists from the beginning.

Bridgman states the central proposition of the operational approach thus : [1]

> " We must demand that the set of operations equivalent to any concept be a unique set, for otherwise there are possibilities of ambiguity in practical applications which we cannot admit. . . .
>
> " *If we have more than one set of operations we have more than one concept*, and strictly there should be a separate name to correspond to each different set of operations."

The word " should " above should be interpreted as meaning that it is *useful*, in terms of certain specific purposes, to adopt the proposed point of view (this applies as well to the word " should " in this sentence). Because of incautious phrasing in his early writing, it has often been thought that Bridgman regarded his own definitions and classifications as logical imperatives. Actually (as

[1] Percy W. Bridgman, *The Logic of Modern Physics* (New York, 1927). The first sentence is on p. 6, the second on p. 10 (my italics).

Although such notions as the meaningfulness and restrictiveness of hypotheses have been made familiar to many economists by followers of Bridgman, the above proposition and the following ones, which are the very heart of the operational approach, are not widely known among economists.

he has since made explicit), it is not necessary to insist on his approach dogmatically or exclusively; without making any unique claims, it is easy to show the value of his point of view (which admittedly is not the most natural) in helping to avoid certain types of confusion.

Of course, in everyday usage we very commonly use the same term to cover different operations, on the grounds that they measure the "same thing." If we take the strict operational point of view that a "thing" is "what is measured by a particular operation," we need not ban the practice of treating two different operations as measuring the "same thing," but we must insist that it be justified by a direct argument. In an important passage, Bridgman indicates the nature of an adequate justification: [1]

> "If we deal with phenomena outside the domain in which we originally defined our concepts, we may find physical hindrances to performing the operations of the original definition, so that the original operations have to be replaced by others. These new operations are, of course, to be chosen so that they give, within experimental error, the same numerical results in the domain in which the two sets of operations may be both applied; but we must recognize in principle that in changing the operations we have really changed the concept. . . . The practical justification for retaining the same name is that within our present experimental limits a numerical difference between the results of the two sorts of operations has not been detected."

It would hardly be possible to find a passage more pertinent to a comparison of the economic theories discussed here. It may be helpful to give some economic illustrations; several examples of pairs of operations which differ but are usually treated as equivalent exist within the boundaries of our discussion.

(1) In the indifference map experiment the operations (*a*) of interrogating the individual as to his preferences, or (*b*) observing his actual choices (Samuelson's "revealed preference"), are usually regarded as alternative.

(2) In the "Jevonsian" experiment two operations are usually thought to be involved: (*a*) inquiring of the subject how he ranks his preferences; (*b*) the subject's own subjective process of "weighing" satisfactions. The first is said to measure differences in satisfaction on the assumption that it approximates the results of the second. The basis of this assumption is that in the area

[1] Percy W. Bridgman, *The Logic of Modern Physics* (New York, 1927); the first two sentences are on p. 23; the third, on p. 16 (in the latter sentence, Bridgman refers to the measurement of length by ordinary and by Einstein's operations).

where they can both be applied—the area of our own introspection—they give identical results (to the extent that we *can* balance satisfactions and that we tell the truth).

(3) In the von Neumann–Morgenstern experiment we used the operation (*a*) of asking the subject to name a p at which he would be indifferent; but we also suggested the possibility (*b*) of observing his choices when confronted with various pairs of prospects many times. Mosteller used the latter operation in his empirical tests.

In each case, the alternative operations are regarded as roughly identical. Actually, most economists who have had practical experience in applied theory are well aware that the results of interrogation and of observing actual behavior are almost never identical. Moreover, minor differences in the operations (such as the wording of questions) do " make a difference." [1] When a pair of operations is accepted as measuring the " same thing " it is because the divergence between the results is not regarded as significant. But as the range of application of each operation is widened over time, divergences appear in the area of overlap, and as precision increases, small differences become significant. Too often, theorists are unprepared for these phenomena and are thrown into confusion at the emergence of ambiguity and paradox.[2] One who accepts the propositions of the operational approach, on the other hand, not only expects these problems to arise but also knows where to watch for them. This (and nothing more pretentious) is the chief virtue which is claimed for the approach.

The relevance of the above discussion to the present problem can now be stated. Probably many readers of *The Theory of Games* and some later articles have received the impression that the third pair of operations above was being proposed as " measuring the same thing " as the second pair. In other words, many have interpreted the von Neumann–Morgenstern experiment as a more precise or practical, though indirect, approach to the results of the Jevonsian experiment: *i.e.*, basically, to the results of the subjective calculation of satisfactions. But if the operational point of view were more common as a habit of thought, readers would have placed the burden of proof on the (supposed) exponents of such an equivalence, challenging them to exhibit

[1] The operational approach may be useful in reminding us that differently worded questionnaires measure, in general, " different things."

[2] Such confusion was prevalent in physics prior to the revolutionary theories of Einstein and Planck. The operational approach was proposed as a means of avoiding such a state of mind in the future.

evidence. In fact, as they would have discovered, there are no such exponents. And the evidence does not exist, for in general the two operations do not produce even approximately the same results.

Let us recall the results of our hypothetical von Neumann–Morgenstern experiment: $U_0 = 0$, $U_{200} = 2$, $U_{500} = 8$, $U_{1,000} = 10$. If the von Neumann–Morgenstern index for this individual were plotted as a function of money incomes, interpolating a smooth curve, the graph would be concave upward between $0 and $500; in this range it would show "increasing marginal utility." This shape would reflect merely the fact that in this range of money outcomes the individual accepted "unfair" bets and that his choices among prospects showed consistency of a certain type.

In contrast to this, Marshall and Jevons predicted almost unconditionally one general feature of a "utility" curve derived from an experiment of the Jevonsian type; it would be concave downwards throughout its whole length, exhibiting non-increasing marginal utility at all points. Among those economists who believe that a Jevonsian experiment can have consistent results at all, few have ever disputed this opinion.

If Marshall's prediction does hold, then the numbers inferred from the two experiments will certainly conflict for any person who is observed to accept a gamble at odds which are not distinctly favorable, let alone odds that are actually unfair. If such a person has a von Neumann–Morgenstern index it will have a range of increasing marginal utility, which is assumed to be contradictory to the Jevonsian index. Marshall himself pointed out that there were such people, even among those otherwise "rational."

Thus we can state: the von Neumann–Morgenstern and Jevons–Marshall operations do *not* measure the "same thing." The former do not simply tend to measure the Marshallian "utilities" with greater precision, *i.e.*, to a higher number of significant figures. In general, the ranking of first differences in "utility" as a function of money will be different, depending on which "utility" is being measured; if the functions are continuous, the second derivatives will not in general have the same sign.

To those who accepted the apparent inference that the gambling operation allowed an "estimate" of the results of a Jevonsian experiment, a moment's thought should have suggested the question: Why is an estimate necessary? If risk-behavior

reveals something about differences in satisfaction, presumably it is because those differences in satisfaction are the decisive factor in decision-making. But in that case we might as well ask about satisfactions directly.[1]

This discussion has not established yet that von Neumann and Morgenstern do not themselves regard their operation, mistakenly, as measuring differences in satisfaction. The evidence for this is their repeated rejection of the notion that an individual reaches decisions in risk-situations by calculating differences in utilities, their brand or any other (such as Jevonsian utilities). But much confusion probably stems from the fact that they are prone to write in large, clear type about comparing differences in preferences and to discard such notions in fine print at the bottom of the page. Thus, they formulate their "continuity" axiom (which rules out the "absolute" rejection of lottery tickets or the "absolute" love of gambling discussed earlier) as follows: [2]

> "No matter how much the utility of v exceeds . . . the utility of u, and no matter how little the utility of w exceeds . . . the utility of u, if v is admixed to u with a sufficiently small numerical probability, the difference that this admixture makes from u will be less than the difference of w from u."

This leaves a strong impression, to put it mildly, that the notions of quantity of utility and differences in quantities are an integral part of the argument . . . unless the reader follows a footnote on to the next page: [3]

[1] It is perhaps conceivable, though unlikely, that his feelings of satisfaction might be difficult for an individual directly, being only semi-conscious, though influencing his behavior. But if it were true that his risk-behavior was a reliable and convenient guide to his feelings of differences in satisfaction, this would ensure that the Jevonsian experiment *could* always be performed. For even if the subject were an economist, say, who detested introspection, he could note tacitly his reactions to hypothetical lottery tickets (or even, if conscientious, plot his behavior at bingo games and horse races) before replying to questions about differences in satisfaction.

On the other hand, to say that the Jevonsian experiment cannot lead to consistent results is to say that any consistency revealed by the von Neumann–Morgenstern experiment is not closely related to satisfaction.

[2] Von Neumann and Morgenstern, *op. cit.*, p. 630. Since the content of this passage is not under discussion, the reader is advised to pass his eyes over it rather swiftly.

[3] *Ibid.*, p. 631 n. This is not the only time in their book that the authors introduce notions in a "literary" discussion of their theorems that they simultaneously disown, informing the reader that it all comes out in the axioms. Of course, the very inclusion of a verbal discussion is a concession to non-mathematicians; but one can do only so much in the name of "heuristic devices." It is not a recommendation of the empirical relevance of axioms to say that they can be made plausible in literary translation only by identifying them with notions (such as subjective utility differences) which are actually irrelevant.

> " The reader will also note that we are talking of entities like ' the excess of *v* over *u*,' or ' the excess of *u* over *v* ' or (to combine the two former) the ' discrepancy of *u* and *v* ' (*u*, *v*, being utilities) merely to facilitate the verbal discussion—they are not part of our rigorous, axiomatic system."

One other passage in the " literary " discussion is probably the greatest single source of misunderstanding; it concerns a situation in which an individual is offered a choice between a sure outcome, *A*, and a 50–50 chance of *B* or *C*, where *C* is preferred to *A* and *A* to *B* [1]:

> " any assertion about his preference of *A* against the combination contains fundamentally new information. Specifically: If he now prefers *A* to the 50–50 combination of *B* and *C*, this provides a plausible base for the numerical estimate that his preference of *A* over *B* is in excess of his preference of *C* over *A*."

This passage seems clearly to imply that the von Neumann–Morgenstern operation aims at the same " entities " (*i.e.*, utility differences) as the Jevonsian experiment, being merely more indirect. But again, the crucial withdrawal is in the footnote : [2]

> " Observe that we have only postulated an individual intuition which permits decision as to which of the two ' events ' is preferable. But we have not directly postulated any intuitive estimate of the relative sizes of two preferences —*i.e.*, in the subsequent terminology, of two differences of utilities."

The equivocal word here is " estimate." This implies that the procedure tries to approximate the results of an introspective operation. Actually, in the von Neumann–Morgenstern experiment described above utility differences were not " estimated " but computed exactly. They were related precisely to certain risk-choices; no other evidence, intuitive or otherwise, was allowed to influence the results.

The authors themselves point out the ambiguity :

> " Are we not postulating here—or taking it for granted—that one preference may exceed another, *i.e.*, that such statements convey a meaning? Such a view would be a complete misunderstanding of our procedure." [3]

Their procedure is actually to use the risk-choices to *define* the utility differences—to make this notion meaningful in a new way —not to " estimate them." Very likely it was the above passage which led Professor D. H. Robertson to imagine that von Neumann

[1] *Op. cit.*, p. 18. [2] *Ibid.*, p. 18 n. [3] *Ibid.*, p. 20.

and Morgenstern had proposed a method for estimating relative differences in desirability. It is easy to spot this inference in his critical account of their theory : [1]

> "Thus in the case of a man who does not know how to choose—*i.e.*, who chooses by the toss of a mental coin—between the certainty of *B* and an even chance of *A* or *C*, these authors offer 1 as a measure of the ratio of *AB* to *BC*. . . . But it is clear that this would only be *true* for a particular type of man, namely, one who is content to be governed entirely by mathematical expectations. . . ." (My italics.)

In the case of the behavior described, von Neumann and Morgenstern would *define* the ratio of utility differences as 1 ; and in a matter of definition there can be no question of truth and falsity. Those standards could be applied only to a *hypothesis* that the scale defined by von Neumann and Morgenstern bore some empirical relation to some other data, not involving risk : for example, the hypothesis that it approximated the results of a Jevonsian measurement. It is clear from the context of Robertson's remarks that he believed, like most readers, such a hypothesis was implied. It has been the argument of this paper that this belief is mistaken.

In the same passage Robertson adopts essentially the Marshallian position : [2] ". . . we can make no sense of his actions in the face of uncertainty without supposing that he can form some estimate of the relative difference in desirability between pairs of situations." Whatever the plausibility of this argument, it has no relevance to the von Neumann–Morgenstern theory. Where Marshall postulated a type of "consistency" between men's risk-choices and their feelings of relative differences in desirability of the outcomes, von Neumann and Morgenstern hypothesise simply a consistency between risk-choices and other risk-choices. By coincidence, it happens that the particular form of "consistency" prescribed by Marshall (he might well have chosen some other rule than the maximisation of "expected utility") would imply von Neumann–Morgenstern "consistency"; though not vice versa. A man who had a Jevonsian index *and* who obeyed Marshall's dictum would have a von Neumann–Morgenstern index; but the existence of the latter index implies neither of the first two conditions (and the existence of the former index implies neither of the last two conditions).

[1] D. H. Robertson, *Utility and all That* (London, 1952), p. 28. [2] *Ibid.*, p. 28.

Thus the von Neumann–Morgenstern axioms cover all those who are "rational" in the Marshallian sense; in addition, they may apply to others who would be "irrational" in Marshall's terms, *e.g.*, bettors who accepted unfair bets [1]: and still others for whom no Jevonsian index can be defined.

Von Neumann and Morgenstern describe their procedure thus: "We have practically defined numerical utility as being that thing for which a calculus of mathematical expectations legitimate." [2] The word "practically" is unnecessary; from an operational point of view, they *have* so defined it. Does such a "thing" exist? Friedman and Savage have emphasised the use of the axioms, which put definite restrictions on behavior, as a basis for testable and fairly powerful predictions concerning risk-choices.[3] *Should* it exist? Marschak has proposed that the axioms be regarded as defining "rational" behavior in risk-situations; according to this view, which no other writer has supported, the axioms are of interest for normative purposes, even if no one actually does conform to them.[4]

Von Neumann and Morgenstern cited only the descriptive aspect of the theory. They were not particularly interested in predicting or prescribing people's preferences among prospects, but merely in describing them in terms of mathematical expectation: a necessity in their own theory of games, a convenience in any context. This original view of the subject, by far the least pretentious, is probably the most appropriate. The emphasis by Friedman and Savage on the meaningfulness of the hypothesis obscures the fact that many other hypotheses are just as meaningful, perhaps more useful, and even more convenient for predictive purposes (though not for description). For example, hypotheses in terms of parameters of the *money* distribution, such as mathematical expectation and variance, might produce fully as good predictions as those based on a derived von Neumann–Morgenstern index, and they would certainly be easier to test. As for the normative aspect, there seems very little reason to advise a man who is extremely reckless (or excessively conservative) in

[1] The theory, since it allows for this sort of behavior, cannot be said to rule out all forms of "pleasure in gambling." But proponents have rather overplayed this point. Although acceptance of unfair bets does not contradict the theory, there is ample behavior which does, including some other forms of "pleasure in gambling."

[2] *Op. cit.*, p. 28.

[3] Friedman and Savage, *op. cit.*

[4] Marschak, *op. cit.*, p. 139; also, "Why 'Should' Statisticians and Businessmen Maximise 'Moral Expectation'?" *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Los Angeles, 1951), p. 493.

some of his risk-choices that he should be consistently reckless (or conservative) in his remaining risk-choices. Nor does it seem that a person who behaves approximately in accordance with a von Neumann–Morgenstern index would be in any sense better off if he behaved *more* in accordance with it. If these conclusions are accepted (and von Neumann and Morgenstern would probably accept them), then one must answer the question, " What does it matter whether such a ' thing ' exists ? " very conservatively.

At any rate, it should be clear that Baumol's impression that " Neumann and Morgenstern consider the utility index obtained by them as the *only* true one " [1] is quite mistaken. So far as behavior under certainty is concerned, only the ordinal features of the index are relevant. The only numerical operation permitted is that of forming mathematical expectations, which is related to risk-behavior; it makes no sense, for example, to *add* von Neumann–Morgenstern utilities. The cardinal features of the index—the relative differences between utility numbers—are used only to predict or describe risk-behavior, and, moreover, are derived solely from risk-behavior. Therefore the results of a von Neumann–Morgenstern experiment cannot be " checked " against the results of any experiment not involving risk-choices. This applies to simple introspection, to the Jevonsian experiment, and also to other attempts to base a cardinal utility on consumer behavior.[2] These latter have been rather thoroughly discredited by Samuelson and others because of their use of special unrealistic assumptions. But the existence or non-existence of a von Neumann–Morgenstern index and the existence of " measurable " indices based on these other operations are entirely independent matters. Each method might, out of the whole set of ordinal utility indices, select a different subset of indices reflecting some type of data in addition to preferences; if the indices inside each subset were related by linear transformations, each method would result in *a* " measurable " utility index. These indices might have entirely different shapes, but so long as they did not entirely overlap in application, there would be no need to single out any one of them as being the " true " utility index. Certainly von Neumann and Morgenstern make no such claims for their construction; they cite only its convenience in

[1] Baumol, " The Neumann-Morgenstern Utility Index—an Ordinalist View," *Journal of Political Economy*, February 1951, p. 61.

[2] See Robert Bishop, " Consumer's Surplus and Cardinal Utility," *Quarterly Journal of Economics*, May 1943. For criticism of these approaches see Samuelson, *Foundations of Economic Analysis*, pp. 174–9.

formalising risk-behavior. There is no reason to believe that a " measurable utility " derived by some other method could do this; [1] on the other hand, the von Neumann–Morgenstern index could not do the main jobs for which other constructions are intended. It would be of no aid whatsoever in formalising consumer behavior under certainty (the goal of the Fisher–Frisch constructions : see Bishop, *op. cit.*), nor would it seem to be of any relevance in welfare evaluations (whereas a Jevonsian index might be).[2] If it is true, as Professor Robertson has complained, that von Neumann and Morgenstern have actually done harm to " the cause " of creating acceptance for a measurable utility with these last two objectives,[3] this is but a measure of the general misinterpretation of their results : a confusion for which they cannot evade all responsibility.

D. Ellsberg

Cambridge, England.
Cambridge, Mass.

[1] Thus, Alchian is mistaken in asserting that " measurability ' up to a linear transform ' both *implies* and is implied by the possibility of predicting choices among uncertain prospects, the universal situation " (" The Meaning of Utility Measurement," p. 49 (my italics)). Actually, it is easy to conceive a " measurable " utility index which is neither derived from nor used to predict risk-behavior.

[2] After I had reached these conclusions, I had the great benefit of conversation with Professor Oskar Morgenstern, who was kind enough to read and discuss with me an earlier version of this paper. Professor Morgenstern confirmed what were then speculations on the implications of the theory; he particularly confirmed that he and von Neumann had envisioned only limited application, to risk-behavior alone.

[3] Robertson, *op. cit.*, p. 28.