Probability and the Attempts to Measure Utility

Paul A. Samuelson

Massachusetts Institute of
Technology Cambridge, Massachusetts

1. By the 1890's Pareto and Irving Fisher came to realize that the *cardinal* measurability of utility postulated by earlier economic theorists such as Bentham and Edgeworth could be dispensed with in favor of purely *ordinal* utility dealing only with more or less. To the modern theorist cardinal utility is irrelevant both for the positive explanation of demand behavior and for normative welfare economics.

Nonetheless, a few still engage in the parlor-game of devising assumptions which define a unique measure of cardinal utility. This can be done by a variety of different tricks, with the most common ones involving an arbitrary assumption of *additive* utility functions. One of the oldest of these procedures has recently come into vogue again; it involves an attempt to identify a utility function by observing the response of the consumer to probability situations. Only in special empirical cases is this procedure valid; and even in the narrow class of cases where not invalid, it is only of limited theoretical interest except as a convenient way of unifying the description of the consumer's *ordinal* behavior with respect to gambling and insurance. In the following I have stated my views on this matter rather dogmatically so as to provide a broad target for discussion and criticism.

2. Bernoulli and Marshall[1] argued that if utility grows linearly with income (marginal utility being constant) then people will risk a 50 per cent chance of a \$1 loss if compensated by an equal chance of a \$1 gain. On the other hand, if the law of diminishing marginal utility holds, it will be necessary to compensate people for such a risk of loss by the chance of a larger gain. Peoples' reactions to gambling can thus not only reveal the qualitative behavior of marginal utility, but the exact *quantitative* properties of the utility function as well.[2]

1) A. Marshall, *Principles of Economics*, 8th ed., pp. 135, 842-3. Also see M. Friedman and L. J. Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, Vol. 54 (1948) pp. 279-304 for valuable discussion and further references. J. v. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, (Prinston, 1944 and 1947) has given the theory a rigorous axiomatic basis. Neither Daniel Bernoulli nor Marshall can be credited with particularly original contributions, but their names have become associated with their theory.

2) Except for scale and origin constants, a determinate function $U=f(x)=\int_0^x f'(s)ds$ can be defined by an infinity of *different* experimental set-ups. E.g. around any income level x_0, we can observe the empirical relation between a 1/2 chance of a loss of h and the 1/2 chance of a compensatory gain, g; this function $g=g(h;x_0)$ determines $f'(x)$ in all but scale. Or for any two income levels, x_1 and x_2 surrounding x_0, there will be an observable unique probability of x_1, p, that will leave the man indifferent as compared to a sure and certain x_0; call this observed function, $p(x_1, x_2; x_0)$. From it too we can determine $f'(x)$. A third method is to study the limiting behavior of $2[g(h;x_0)-h]/h^2$ as h goes to 0. This defines $\partial^2 g/\partial h^2$, which can be shown to equal $-2f''(x_0)/f'(x_0)$; and from this we can easily get the elasticity of the marginal utility function and (by integration over x_0) its complete shape. An infinity of other similar experiments can be devised; and except in an unlikely special empirical case, each method can be expected to yield a *different* utility function.

Some of the gross facts about gambling for some people are consistent with this theory—e.g. the classical case of a man who never gambles at mathematically fair odds and who pays to be hedged by insurance. But the perfectly possible case of a man who refuses fair small bets at all income levels and yet buys lottery tickets can be handled only by going beyond this simple theory. As yet I know of no empirical predictions that this theory has suggested which have turned out to be (1) valid and (2) novel or inexplicable without this special theory. I may also record the personal view that the sociology of gambling is infinitely richer than this particular theory permits: There is as much to be learned about gambling from Dostoyevsky as from Pascal.

3. My present purpose, however, is not to examine the factual basis of the Bernoulli-Marshall theory of numerical utility, but to show its arbitrariness at the logical level. Moreover, I shall not criticise its gravest defect of bypassing the basic philosophical problem of induction: No philosopher has yet provided the bridge between purely deductive mathematical probability (combinatorial analysis, set and measure theory) and the empirical problem of making a finite number of decisions. Mathematicians, properly, ignore this problem and confine themselves with defining procedures which will be optimal if followed an infinite number of times under ideally defined conditions. But whether I should risk my only child's life in an operation, or believe a witness in court, or back a certain horse or investment project—on these questions mathematical probability gives little counsel.

This basic problem is not peculiar to the Bernoulli-Marshall theory, so I too shall bypass it here and grant that probabilities p_1, $p_2\cdots$ of income levels $x_1, x_2\cdots$, have a meaning and relevance to the individual's single decisions. More specifically, given any two situations

$$\text{A} \quad x_1^a, x_2^a\cdots;\ p_1^a, p_2^a\cdots$$
$$\text{B} \quad x_1^b, x_2^b\cdots;\ p_1^b, p_2^b\cdots$$

with $\sum p=1$, I assume the individual can always decide whether A is worse than B, or B is worse than A, or A and B are indifferent.[3)]

In effect, this means that we can define an indefinite number of numerical indexes or indicators of ordinal preference, of the form

$$V(x_1, x_2, \cdots;\ p_1, p_2, \cdots)$$
$$\text{or} \quad W(V)=W(x_1, x_2, \cdots;\ p_1, p_2, \cdots)$$

with A worse than B implying $V(\text{A})<V(\text{B})$, A indifferent to B implying $V(\text{A})=V(\text{B})$, and with $W(V)$ any one-directional function or renumbering. The indifference loci determined by V or W held constant are empirically identifiable by behavioristic experiments.

For the sake of keeping the exposition simple we may postulate continuity and differentiability of the functions. What properties can we expect of this ordinal preference pattern showing reactions to risk? Unless we confine ourselves to relatively "rational" men, very few *a priori* restrictions can be placed on the data. And if we do agree to confine ourselves to "rational" men, then there is danger of ending up with completely tautological semantic results that entirely depend upon what we choose to read into the word "rationality." Thus, we might end up with the fatuity: "Rational men fulfill the Bernoulli-Marshall conditions, because that is the definition of rationality."

3) Moreover, I assume that this has all the properties of consistent ordinal preference: e.g. A worse than B, and B worse than C, means A worse than C, and similar transitivity relations.

Actually, there is no need to descend to this level of inanity. Certain "reasonable" properties of the V or W functions can be hypothesized.* For example, in the simple case of only two income situations, where $p_1=1-p_2=p$, our indifference contours are defined by surfaces

$$W[V(x_1, x_2; p)]=\text{constant}$$

in the (x_1, x_2, p) space. There is no harm in confining ourselves to the region in which $x_1 \geqq x_2$, so that we can always think of p as the probability of the *larger* of the two incomes, with $1-p$ the other probability.

Then the obviously reasonable requirements on the preference pattern are as follows:

Basis Hypothesis: An increase in any of the three variables, x_1 or x_2, or p, will tend to lead to higher preference.

Thus, raising one of the incomes and changing nothing else will certainly give the man a new situation that has everything the previous situation had, and *something additional* too. Or increasing the probability of a larger income at the expense of a smaller one should certainly make him better off. Only in the limiting case where the two incomes are equal will changes in p be of complete indifference.

The task of the empirical statistician is to record for the human guinea-pig the exact form of this one-parameter family of indifference-surfaces. Aside from the Basic Hypothesis he has no legitimate right to except these surfaces to satisfy any special laws. Needless to say he has no right to expect that the behavior of the surfaces in one part of the space dictates how they *must* behave in any other part of the space—any more than we have a right to extrapolate from a poor man's consumption of tea what his consumption of yachts *must* be when he is rich.

4. The Bernoulli-Marshall theory can be easily shown to involve not-easily-recognized special arbitrary assumption about the family of indifference surfaces—namely that all of the surfaces *everywhere* can be determined by heroic extrapolation, from the behavior of their partial derivatives upon an *arbitrary curve in space.*[4]

In my judgment this is nonsense. The most rational man I ever met, whom I shall

* Even this hypothesis appears to me to be inconsistent with the rich sociology of gambling and risk-taking.

4) The earlier footnote dealing with three out of an infinity of ways of identifying $V=\int_0^x f'(s)\,ds$ provides examples. A geometrical way of stating the Bernoulli-Marshall straight-jacket is as follows: for fixed $(x_1^a, x_2^a; p^a)$, the relation $W(x_1, x_2; p^a)$ $W(x_1^a, x_2^a; p^a)$ defines a horizontal indifference curve in space. There will always exist a mutual stretching of the axes, $x_i=f(x_i)$ which will make *this* curve a straight line. But only in the Bernolli-Marshall curve will this stretching make *all the surfaces satisfy arbitrary relations* $px_1+(1-p)x_2=$constant, with all horizontal indifference curves becoming perfectly straight lines.

Necessary and sufficient conditions for the empirical data to be of the B-M form can be written in many equivalent ways. One operationally meaningful procedure is to define.

$$G=\log(1-p)-\log p+\log(-\partial x_2/\partial x_1) \quad v=\text{constant}$$

Then

$$\frac{\partial G}{\partial p}\equiv 0, \quad \frac{\partial^2 G}{\partial x_1 \partial x_2}\equiv 0, \quad G(x_1, x_2)\equiv -G(x_2 x_1)$$

are together necessary and sufficient conditions; of course,

$$G(x_1, x_2)=\log f'(x_2)=\log f'(x_1)$$
$$=\int_{x_1}^{x_2}[d\log f'(x)/dx]dx.$$

Prof. Jacob Marschak of Chicago has worked out some further conditions that must be satisfied when there are more than two income situations. My colleague, Prof. Robert L. Bishop, has worked out a variety of consistency conditions that follow from B-M theory. I have also been informed by Prof. William J. Baumol of Princeton that he has unpublished criticisms of the B-M theory.

call Ysidro, determined his own ordinal preference pattern and found that it satisfied the exact equation of the well known "ideal index"

$$W = W[(p_1\psi_1 + p_2\psi_2)^{\frac{1}{2}}(p_1\psi_1^{-1} + p_2\psi_2^{-1})^{-\frac{1}{2}}]$$

where $p_1 + p_2 = 1$, and $W'(V)$ is an arbitrary positive function.[5] When told that he did not satisfy all of the v. Newmann-Morgenstern axioms,[6] he replied that he thought it more rational to satisfy his preferences and let the axioms satisfy themselves. Once the empirical implications of the v. Neumann-Morgenstern axioms are understood, their arbitrariness and that of the Bernoulli-Marshall theory stands revealed.

The history of statistical theory is replete with cases where writers have postulated innocent-seeming restrictions and achieved far-reaching and arbitrary results. The great Gauss himself provided two examples which he came later to regret: He once thought an ideal "mean" should be a symmetric and continuously differentiable function which (1) grows by a if each observation grow by a, ahd which (2) is multiplied by a scale factor, b, when each observation is so multiplied. This leads to the arithmetic mean as the ideal statistic, a highly arbitrary result that stayed in many text-books for a century. Similarly, he made an error in calculating a maximum likelihood statistic and ended up, in effect, defining the "ideal curve of error" as that function for which the arithmetic mean is a maximum-likelihood statistic. He might have achieved the same gratuitous result as cheaply by defining the "perfect" error function as that one for which $\sum x/n$ and $\sum(x-\sum x/n)/n$ are independently distributed. Some text-book writers have also followed Clerk Maxwell's similar-type proof of the normal curve based on arbitrarily postulated invariances under axis-rotations of coordinates. Eddington has recently carried this ancient art of *a priori* reasoning even further. If only the data knew what men know, planets might move in perfect circles and incomes might be distributed along Pareto curves.

5. Why did such a plausible theory of utility maximizing lead to such implausible—if not down-right nonsensical—results? I suppose the answer lies in the fact that it is not really a very plausible theory once you examine it carefully. Those who are familiar with the magic by which Irving Fisher, Frisch, and Samuelson were able to put the rabbit of cardinal utility into their hats[7]

5) Ysidro's father and mother had B-W functions, but he inherited a blend of them which is *not* such a function.

6) J. v. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 2nd ed., 1947, pp. 17-31, and Appendix. See my appendix for more details. In the future I hope to give a more lengthy survey of these axioms.

7) Irving Fisher, "A Statistical Method for Measuring 'Marginal Utility' and Testing the Justice of a progressive Income Tax" in *Economic Essays in Honor of John Bates Clark* (1927). R. Frisch, "New methods of Measuring Marginal Utility," *Beiträge zur ökonomischen Theorie*, no. 3 (1932). These involve an "additive" assumption in the field of budgetary consumption data. Still another type of additive assumption in utilities over time—along the lines of Böhm-Bawerk's first ground for interest because of expected increases of future income—permits a unique measurement of conventionally-defined cardinal utility; see P. A. Samuelson, "A Note on the Measurement of Utility," *Review of Economic Studies*, Vol. 4, (1937), pp. 155-61, where some, but not all, of the empirical implications of this procedure are indicated. A not-at-all-obvious "independence assumption" can be shown to give v. Neumann and Morgenstern their results. Readers interested in cardinality of utility will find technical writings of Bishop, Vickrey, Lange, Bernardelli, Lerner, Armstrong, Zeuthen, *et al* of interest. Cf. K. Menger "Das Unsicherheitsmonent in der Wertlehre," *Zeitschrift* für Nationalokönomie, Vol. 5 (1934) pp. 458-85, *especially* 481 ff.

easily recognize that the use of an *arithmetic mean* involves an arbitrary *additive* assumption. Why not the median? Or some other type of mean? Many of these, but not all, would negate the Bernoulli-Marshall theory.

I think the acceptance of "mathematical expectation of utility" or its "arithmetic mean" was an unthinking carryover from the mathematical theory of the law of large numbers as applied to asymptotic processes. Suppose two gamblers each have an infinite amount of money or credit and they gamble together at "potentially fair odds" and infinite number of times, or a very large number of times; suppose one of them acts to maximize the arithmetic mean of his money winnings (*not* of their utility:) and the other maximizes something different. Then as the length of play grows, the probability *approaches in the limit* unity that the first man's winnings will exceed any prescribed number. For finite sequences, however long they may be, the basic philosophical ploblem remains; and even for infinite sequences, the theory seems already to have assumed away any change in marginal utility by its assumption of infinite wealth. Where money is concerned, the additive assumption of the arithmetic mean has an inherent rationals, because *coins are added to coins* in forming a stock of wealth. But to assume that there is *a utility bank in which people make deposits and withdrawals over time* is not only implausible nonsense, but comes close to begging the issue.

A possible limited valid use of the Bernoulli-Marshall methods is in the smoothing of imperfectly observed parts of the surfaces; this is quite different from using them for extrapolation to unobserved parts of the space. Unfortunately, existing experimental techniques seem too crude to make the most interesting tests of all—namely, the extent to which the Bernoulli-Marshall specializations are invalid or valid.*

* And even if there should turn out to be a non-empty class of people for whom these special relations hold, we must not forget that it is their *ordinal* behavior that is of interest; there is no special significance to be attached to the convention of calling the special index of utility $V=pf(x_1)+(1-p)f(x_2)$ the true measure of utility: $W(V)=f^{-1}(V)$ has the interesting property $W(x, x; p)=x$, but it too has no privileged status as numbering of ordinal utility.

APPENDIX

I suspect that I must be quite confused in my interpretation of the logical basis of the v. Neumann-Morgenstern and Friedman-Savage theories since so many eminent mathematicians and economists rarely go wrong in the field of pure deduction. I am a little fearful, therefore, to confess that I regard both systems to be unacceptable, and as far as I can see not even consistent between themselves.

Friedman and Savage (*op. cit* pp. 287-8) base their whole logical case on the following:

[α] "The hypothesis that is proposed for rationalizing the behavior just summarized can be stated compactly as follows: In choosing among alternatives open to it, whether or not these alternatives involve risk, a consumer unit (generally a family, sometimes an individual) behaves as if (a) it had a consistent set of preferences; (b) these preferences could be completely described by a function attaching a numerical value—to be designated "utility"—to alternatives

each of which is regarded as certain; (c) its objective were to make its expected utility as large as possible. [β] It is the contribution of von Neumann and Morgenstern to have shown that an alternative statement of the same hypothesis is: An individual chooses in accordance with a system of preferences which has the following properties:

1. The system is complete and consistent; that is an individual can tell which of two objects he prefers or whether he is indifferent between them, and if he does not prefer C to B and does not prefer B to A, then he does not prefer C to A. (In this context, the word 'object' includes combinations of objects with stated probabilities; for example, if A and B are objects, a 40-60 chance of A or B is also an object.)

2. Any object which is a combination of other objects with stated probabilities is never preferred to every one of these other objects, nor is every one of them ever preferred to the combination.

3. If the object A is prefered to the object B and B to the object C, there will be some probability combination of A and C such that the individual is indifferent between it and B" [Two footnotes omitted by me.]

To me [α] is completely arbitrary and inadmissable, while [β] appears quite acceptable and rather harmless. Indeed my continuity assumptions plus the Basic Hypothesis on the V and $W(V)$ functions, which I shall call [γ], seem to me to be equivalent to [β], except for a few technical details concerning the overstringency of my differentiability conditions. Ysidro's function I believe satisfies [β] but not [α].

Yet Friedman and Savage believe that v. Neumann and Morgenstern have shown the complete equivalence of [α] and [β]. The complete axioms of v. Neumann and Morgenstern (*Theory of Games*, pp. 26-7) are too long to quote here; let us call them [δ]. In the sense in which [δ] is logically equivalent to [α], it must also be unacceptable to me. Therefore, I must doubt that both [β] ≡[δ] and [δ]≡[α] are true.

How can I account for all my strange views? The most likely explanation is that I am simply confused. But assuming the contrary, let me record the suggestion that the v. Neumann-Morgenstern axioms, in the sense that they are equivalent to [β] and in the sense that they are economically acceptable, are *not* equivalent to [α]. How have the authors of the *Theory of Games* deceived themselves as to the inevitability of their demonstration of the measurability of utility—if it should turn out that they are wrong?

My tentative guess is as follows: they have not made a simple error in logic, but have implicitly added a hidden and unacceptable premise to their axioms. The empirical content of their axioms can be translated into the terminology of ordinal utility, $W = W[V(x; p)]$, and they then become unobjectionable. In this purely ordinal context, let us call the axioms [δ]′ rather* than

---

* In terms of their axioms 3: A to 3: C, I think [δ]′ would read

A: There exists $V(x_1, x_2\cdots; p_1, p_2, \cdots)$ and $W(V)$ functions which satisfy the usual transitivity relations.

B: The W and V functions have appropriate continuity properties, and $\partial V/\partial x_i > 0$; $\dfrac{\partial V/\partial x_i}{\partial V/\partial x_j} - 1$ has the sign of $(x_i - x_j)$.

C: V is symmetrical and depends only on the final income sitution, no matter how the lottery tickets and probabilities are compounded.

On this definition [β]≡[γ]≡[δ]′≠[δ]≡[α], as I hope to show in a later paper.

$[\delta]$. I believe there to be a world of difference between $[\delta]$ and $[\delta]'$.

Whatever the logical validity and economic admissibility of the complete v. Neumann-Morgenstern axioms, the preliminary literary discussion leading up to these axioms appears open to objections. For example, the authors follow the excellent example of Pareto; Bowley, and Lange and argue that if we can always *ordinally* relate any two *change in well-being*, then we can define a *cardinal* measure of well-being or utility. Thus, if I cannot go beyond statements of the type: "I like Paris better than London, and I like New York better than Chicago," then only *ordinal* utility statements are possible. But if I can make statements like the following:. "I like Paris as much better than London as I like New York better than Chicago," then a numerical scale of utility can be defined.

Now there are some subtle difficulties with this type of argument and certain implicit assumptions must be made if the result is to follow.* But let us waive these subtleties and for the moment grant the authours this convention that *the cardinal utility to me of Rome is exactly half way between that of London and Paris if I am indifferent between the certain prospect of going to Rome and a* (.5, .5) *chance of going to London or Paris.* This has been shown by my colleague Prof. Robert L. Bishop *not* to define a satisfactory scale of utility in the general case. Only in the very special Bernoulli-Marshall case will a unique self-consistent scale be defined. If the validity of the "Bishop-effect" is granted, then the authors appear in this part of their discussion to have begged the question at issue.

To see why their procedure leads to contradictory scales, suppose that they have found five situations that are equally spaced in their defined metric.

milk *wine* tea *fruit juice* coffee

This means that a certain and sure cup of tea is equally attractive to a (.5, .5) chance of getting a cup of fruit juice or of wine; and so forth for the other items. Now let us exclude the italicized intermediate items so that we have the sequence

milk tea coffee

For a general surface such as is described in this paper, will it be necessarily true that tea is then "half way" between milk and coffee?

The answer is "no, not necessarily," as Prof. Bishop has shown. I hope he will publish his discussion at some later date.

---

* See O. Lange, "The Determinateness of the Utility Function," *Review of Economic Studies*, Vol. 1 (1934), pp. 218-25 and later articles appearing in the same journal over the next five years by R.G. D. Allen, Phelps Brown, Bernardelli, Lange, and Samuelson. F. Alt published in the Zeitschrift für Nationalökonomie (1936) and axiomatic treatment of a similar problem.