A Critique of Expected Utility Theory: Descriptive and Normative Considerations

Author(s): Amos Tversky

Source: *Erkenntnis (1975-)*, Jun., 1975, Vol. 9, No. 2 (Jun., 1975), pp. 163-173

Published by: Springer

Stable URL: https://www.jstor.org/stable/20010465

AMOS TVERSKY

A CRITIQUE OF EXPECTED UTILITY THEORY: DESCRIPTIVE AND NORMATIVE CONSIDERATIONS*

Expected utility theory has dominated the analysis of decision making under uncertainty. The expected utility principle was formulated in the 18th century by Daniel Bernoulli (1738), it was first axiomatized by von Neumann and Morgenstern (1944), and it was further developed by Savage (1954) who integrated the notion of subjective probability into expected utility theory. The theory of expected utility, or utility theory for short, has been used in economics as a descriptive theory to explain various phenomena such as the purchase of insurance and the relation between spending and saving. Utility theory has also been employed as a normative theory, in decision analysis, to determine optimal decisions and policies. The axioms of utility theory (e.g., transitivity, substitutability) are accepted by most students of the field as adequate principles of rational behavior under uncertainty. In contrast, there is considerably less agreement regarding the descriptive validity of these axioms.

The experimental tests of utility theory did not yield unequivocal results. Several studies showed that, under special circumstances, most axioms of utility theory are violated. Conditions under which transitivity, for example, is violated are described in Tversky (1969) and Raiffa (1968, page 75). Other experimental studies, however, tended to support utility theory, see, for example, Mosteller and Nogee (1951), Davidson *et al.* (1957), and Tversky (1967). Thus, it remains unclear whether utility theory provides a reasonable approximation to the behavior of individuals under conditions of uncertainty.

In this lecture, I would like to discuss some new experimental results, obtained by Professor Daniel Kahneman and myself at the Hebrew University in Jerusalem. The results of our studies show that utility theory, under the standard interpretation, is grossly inadequate as a descriptive model of individual choice behavior. Furthermore, the results raise serious questions regarding the interpretation of consequences or outcomes, and highlight some of the difficulties involved in the normative application of utility theory. For the purpose of this lecture, there is no

Erkenntnis **9** (1975) 163–173.

need to describe the experimental procedures and the data in detail. Consequently, I will only describe a few selected problems and indicate the general nature of the results. A detailed report of the data will be presented elsewhere.

Decision making under uncertainty can be viewed as choice between gambles, or lotteries. For simplicity, we shall restrict the discussion to gambles with monetary outcomes and (so-called) objective probabilities, although much of our conclusions also apply to more complicated gambles. In particular, we will be primarily concerned with two-outcome gambles of the form (X, P, Y) where one receives an amount X with probability P and an amount Y with probability $1-P$. Perhaps the most salient finding concerning the choice between gambles is the presence of risk aversion. For example, most people prefer \$400 for sure over the gamble (\$1,000, 1/2, 0) although the expected actuarial value of this gamble is \$500. Similarly, most people are not willing to accept the gamble (\$1,000, 1/2, −\$999) although its expected value is positive.

Indeed, similar considerations have led Daniel Bernoulli in 1738 to replace expected value by expected utility as a criterion for choice between gambles, and to propose that the utility function for money is a concave (i.e., negatively accelerated) function of money. In the modern axiomatic approach to utility theory, the concavity of the utility function is not assumed but rather inferred from the preferences. For modern treatments of risk aversion, see Pratt (1964), and Arrow (1971). (Formally, a person is risk-averse if and only if he prefers any sure-thing X over any gamble with expected value X). In utility theory, risk aversion is explained by the concavity of the utility function for money. Once the monetary scale is properly transformed – no risk aversion remains. (In this respect it is somewhat misleading to refer to the measurement of the utility for money as 'the measurement of attitudes towards risk'. One's utility function reflects one's attitude toward money, not towards risk. Risk aversion is an epiphenomenon in utility theory).

Evidently, the utility for gains (but not necessarily for losses) is negatively accelerated. The subjective difference between the status quo and a gift of 1 million dollars is clearly greater than that between a gift of 1 million dollars and a gift of 2 million dollars. The subjective value of money, like the subjective value of brightness, loudness, or weight, is a concave function of the respective physical measure. However, as we shall demonstrate below, the concavity of the utility of money cannot account for the observed pattern of choices.

Consider the following decision problems. For simplicity, we use (X) to denote the option of receiving $\$X$ for sure.

$$\begin{array}{llll} \text{Situation I:} & A = (1{,}000,\ 1/2,\ 0) & & B = (400) \\ \text{Situation II:} & C \quad (1{,}000,\ 1/10,\ 0) & & D \quad (400,\ 1/5,\ 0) \end{array}$$

The great majority of subjects prefer B over A and C over D. This pattern is highly consistent: it is obtained with different monetary outcomes and different probabilities, and it holds for naive as well as for highly educated subjects. Nevertheless, it is incompatible with utility theory. To demonstrate, note that if B is chosen over A then, under utility theory (with $U(0)=0$), we obtain $U(400)>1/2U(1{,}000)$. On the other hand, if C is chosen over D, we obtain $1/10U(1{,}000)>1/5U(400)$, a contradiction. Thus, in utility theory B is preferred to A if and only if D is preferred to C. Note that C can be expressed as $(A,\ 1/5,\ 0)$, whereas D can be expressed as $(B,\ 1/5,\ 0)$. That is, C and D can be represented as gambles with 1/5 chance to obtain A and B respectively. In general, utility theory requires that whenever one option, say A, is preferred to another option, say, B, then any probability mixture of A with 0 should be preferred to the same probability mixture of B with 0. This is the substitution condition of utility theory which has been flagrantly violated by our subjects.

To test utility theory further, we have presented subjected with another pair of options, labelled E and F, defined by the following two-stage game. In the first stage, you have a probability of 1/5 to move into the second stage, and a probability of 4/5 to terminate the game at the first stage without winning anything. If you move to the second stage you have a choice between two options: in option E you have a 50–50 chance to get \$1,000 or nothing, and in option F you get \$400 for sure. You have to make your choice before the game is played, that is, before the outcome of stage 1 is known. Note that option E can be represented as $(A,\ 1/5,\ 0)$ while option F can be represented as $(B,\ 1/5,\ 0)$. Hence, from the standpoint of utility theory, E and C are the same, and F and D are the same. Nevertheless, in this situation, the majority of subjects chose option F over option E, while in the previous comparison the majority of subjects chose option C over option D.

What conclusions can be drawn from these data? The choice of B over A shows the presence of risk aversion. However, the choice of C over D indicates that the preferences are incompatible with any utility function. In fact, the data reveal a *positive certainty effect,* namely, a preference for (positive) outcomes that are obtained with certainty. Thus, the utility of a positive outcome appears greater when it is certain than when it is embedded in a gamble. According to this interpretation, B is chosen over A not because $2U(400) > U(1{,}000)$ but rather because B has a certainty advantage over A. Indeed, when both outcomes are uncertain, the reverse inequality obtains. The finding that gamble F was chosen over E provides further support for this interpretation. Here, the subjects apparently disregarded the outcome of the first stage (which was the same for both gambles) and compared the options according to the outcomes of the second stage where F enjoys a certainty advantage over E.

So far, we have demonstrated the failure of the substitution condition with gambles whose outcomes were either positive or zero. How does utility theory hold up in choices involving negative outcomes? Let $(-X, P, 0)$ be a gamble in which one loses X dollars with probability P and nothing with probability $1-P$, and let $(-X)$ denote the certain loss of X dollars. Consider now the following decision problems.

$$\begin{array}{lll} \text{Situation I:} & A' = (-1000, 1/2, 0) & B' = (-400) \\ \text{Situation II:} & C' = (-1000, 1/10, 0) & D' = (-400, 1/5, 0). \end{array}$$

Note that these gambles are the mirror images of the gambles described earlier: the two sets differ only in the sign of the outcomes. Of course, all the above gambles are undesirable. However, when asked to choose between them, most subjects prefer A' over B' and D' over C'. Note that the preference ordering between the negative gambles is precisely the opposite of the preference ordering between the positive gambles. Following our earlier analysis, it is readily seen that these preferences are also incompatible with utility theory. The substitution condition of utility theory requires, in this case, that A' is preferred to B' if and only if C' is preferred to D'. The data strongly violate this condition.

As in the positive case, we have presented subjects with another pair of options, labelled E' and F', defined by the following two-stage game. In the first stage, you have a probability of 1/5 to move into the second stage, and a probability of 4/5 to terminate the game at the first stage without losing anything. If you move to the second stage, you have a choice between two options: in option E' you have a 50–50 chance to lose \$1,000 or nothing, and in option F' you lose \$400 for sure. You have to make your choice before the game is played. That is, before the outcome of stage 1 is known. Note that option E' can be represented as $(A', 1/5, 0)$, while option F' can be represented as $(B', 1/5, 0)$. Hence, from the standpoint of utility theory, E' is identical to C', and F' is identical to D'. Nevertheless, a majority of subjects chose option E' over F' in the present problem and option D' over C' in the previous problem.

What conclusions can be drawn from the observed preferences between negative-outcome gambles? The data exhibit risk-seeking. Most subjects preferred the gamble A' to the certain option B' despite its inferior expected value. In utility theory, risk seeking can be accounted for in terms of a convex region in the utility function for money, see e.g., Friedman and Savage (1948). However, the choice of D' over C' shows that the preferences are incompatible with any utility function. In fact, the data reveal a *negative certainty effect*, that is, an aversion for sure losses. Thus, the disutility of a negative outcome appears greater when it is certain than when it is embedded in a gamble. According to this interpretation, A' is chosen over B' not because $U(-1{,}000) > 2U(-400)$, but rather because people wish to avoid the sure loss of \$400. Indeed, when both outcomes are uncertain, the reverse inequality obtains.

Further support for the proposed interpretation is provided by the preference of option E' over F'. Here, the subjects apparently disregarded the first stage of the game (which is the same for both options), and compared the options according to the outcomes of the second stage of the game. In this case, option F' becomes a sure-loss and E' is selected.

In summary, the results provide strong evidence against the substitution condition which underlies the theory of expected utility. In particular, the substitution condition was violated when a sure thing was embedded in a gamble. When all outcomes were non-negative, subjects showed a strong preference for sure-things. When all outcomes were non-positive, the subjects showed a strong aversion to sure-things. These phenomena, called certainty effects, produced (some) risk aversion for positive outcome gambles, and (some) risk seeking for negative outcome gambles. Both effects, however, could not be accommodated by any (concave or convex) utility function for money. To accommodate these data, utility theory should be revised in several important respects. The development of a more adequate descriptive theory of choice, however, lies beyond the scope of the present lecture.

So far, I have argued that utility theory is not adequate as a descriptive theory of choice. This conclusion restricts the applicability of the theory to psychology, economics, and other disciplines that attempt to explain decision making under risk. Utility theory, however, can also be used as a prescriptive or normative theory. Indeed, many decision theorists are not concerned with the descriptive adequacy of utility theory because they regard it as a normative theory whose purpose is to prescribe choices rather than describe them. In discussing violations of utility theory, Raiffa (1968, pp. 81–82) writes:

> The primary reason for the adoption of a prescriptive or normative theory... is the observation that when decision making is left solely to unguided judgment, choices are often made in an internally inconsistent fashion, and this indicates that perhaps a decision maker could do better than he is doing. If people always behaved as this prescriptive theory says they ought to, then there would be no reason to make a fuss about a prescriptive theory. We could then just tell people to 'do what comes naturally'.

To motivate the discussion of the normative force of utility theory, let us examine a well-known counter-example, proposed by the French economist Maurice Allais (1953). Consider the following decision problems.

Situation I:	A	\$1,000,000 for sure
	B	\$1,000,000 with probability 0.89
		\$5,000,000 with probability 0.10
		Nothing with probability 0.01
Situation II:	C	\$1,000,000 with probability 0.11
		Nothing with probability 0.89
	D	\$5,000,000 with probability 0.10
		Nothing with probability 0.90

In these problems, most subjects prefer option A over option B and option D over option C. These preferences again violate utility theory. To demonstrate, suppose $U(0)=0$, then the former preference implies that $0.11\ U(1{,}000{,}000) > 0.10\ U(5{,}000{,}000)$, while the latter preference implies the reverse inequality. Specifically, the predominant choices in Allais' problem violate the independence assumption of utility theory according to which the preference between options should not be affected by changes in outcomes that are common to both options. Note that the pattern of choices obtained in Allais' problem is closely related to the findings reported earlier. In this case, the outcome of obtaining 1 million dollars enjoys the certainty advantage in the first comparison (between option *A* and option *B*) but not in the second comparison (between option *C* and option *D*). Although the previous examples were regarded as violations of the substitution condition while Allais' example is viewed as a violation of the independence axiom, they appear to be governed by the same psychological process – the certainty effect. Aside from the minor difference in structure, Allais' example differs from the earlier ones in that it involves considerably larger sums of money and more extreme probability values.

Allais' example has generated a great deal of discussion among decision theorists. They all admitted that, when first presented with the problem, they were tempted to choose option *A* over option *B* and option *D* over option *C*, contrary to utility theory. Upon further reflection, however, some people revise their preferences in accord with the axioms while others remain faithful to their initial responses. Allais argued that the choice of *A* over *B* and *D* over *C* is not only common and natural but also perfectly rational. According to his view, the axioms of utility theory rather than people's preferences should be revised. Other decision theorists, took a different position. In particular, Savage (1954) has proposed the following analysis of the problem, which led him as well as many others, to revise their preferences in accord with utility theory.

Savage represents the options in Allais' problem as a lottery with 100 numbered tickets. The prizes associated with the tickets are described in Table I (in millions of dollars).

Now, it is readily seen that if we draw a ticket whose number is between 12 and 100, then *A* is identical to *B* and *C* is identical to *D*. Hence, argues

TABLE I

		Ticket Number		
		1	2–11	12–100
Situation I:	*A*	1	1	1
	B	0	5	1
Situation II:	*C*	1	1	0
	D	0	5	0

Savage, we should focus our attention on the possibility that the number of the selected ticket is less than 12. In this case, however, option *A* is identical to option *C* and option *B* is identical to option *D*. Consequently, *A* should be selected over *B* if and only if *C* is selected over *D*. Similar analyses leading to the same conclusion were presented by Raiffa (1968, pp. 80–86).

Savage's analysis purports to show that the common pattern of preferences in Allais' problem violate the independence principle of utility theory. As such, it is undoubtedly illuminating. But does it have normative force? Both Savage and Allais agree that the data violate the independence principle; they disagree on the implications of this result. Allais defends the rationality of the preferences and denounces the normative adequacy of utility theory, while Savage regards the preferences in question as erroneous, and defends utility theory. I beg to disagree with both of them. Specifically, I wish to argue that the key issue is not the normative adequacy of the independence principle, but rather the legitimacy of various interpretations of the outcomes, and that this issue lies outside the scope of utility theory. As a consequence, I wish to disagree with both Savage and Allais on the point on which they agree with each other, and to agree with both of them on points on which they disagree with each other.

Let us examine carefully Savage's analysis of Allais' problems. This analysis is based on the assumption that events which do not distinguish between the relevant options can be discarded. Since any ticket whose number is 12 or more yields an outcome that is independent of one's choice, one should choose as if one knows that the number of the selected ticket will be less than 12. But why should one act in this manner? The force of Savage's analysis is explicative rather than prescriptive. It shows why the common preferences in Allais' problems are incompatible with the standard interpretation of utility theory. The analysis does not show that these preferences are irrational or mutually inconsistent.

Furthermore, Savage assumes that the consequences under study are fully characterized by the relevant monetary values. This assumption cannot be derived from normative considerations, and in any case is not part of utility theory. To illustrate, compare the zero outcomes in options *B* and *D*. It has been argued that the failure to win in option *B* (when one had a chance to win a million dollars for sure) is a less desirable outcome than the failure to win in option *D* (when no sure thing was available).

If people perceive the outcomes in this manner, then the description of the consequences should be properly expanded. For example, the zero outcome in option *B* should be described as 'missing the prospect of getting one million dollars for sure'. But if the effective consequences were to include more than the monetary values then Savage's analysis no longer applies, and utility theory is no longer violated.

The question of whether utility theory is compatible with the data or not, therefore, depends critically on the interpretation of the consequences. Under the narrow interpretation of Allais and Savage, which identifies the consequences with the monetary payoffs, utility theory is violated. Under the broader interpretation of the consequences, which incorporates non-monetary considerations such as regret, utility theory remains intact. The key issue, therefore, is not the adequacy of the axioms, but rather the appropriateness of the interpretation of the consequences.

The theory of expected utility is formulated in terms of an abstract set of consequences, that are the carriers of utilities. The axiomatic theory, by its very nature, leaves the consequences uninterpreted. Any application of the theory, of course, is based on a particular interpretation of the outcomes. Thus, the theory could be valid in one interpretation and invalid in another. The appropriateness of the interpretation, however, cannot be evaluated within the theory. How then should the consequences be interpreted in any particular application?

In the absence of any constraints, the consequences can always be interpreted so as to satisfy the axioms. In this case, however, the theory becomes empty from both descriptive and normative standpoints. In order to maximize the power and the content of the theory, one is tempted to adopt a restricted interpretation such as the identification of outcomes with monetary payoffs. In this case, the normative force of the theory depends critically on the appropriateness of the selected interpretation. Why should one exclude all non-monetary considerations in choosing between gambles? Surely there are many contexts in which considerations of regret are justifiable. Why are they illegitimate in decision making under risk? Utility theory is incapable of answering these questions.

Normative decision theory does not provide a complete analysis of rational decision making in the face of uncertainty. In the applications of decision analysis, one treats the utilities as given, and focuses on the manner in which they are combined. In this respect, normative decision theory is concerned only with the consistency of preferences, not with their justification. Put differently, decision theorists are eager to tell people how to act, in light of their values, but they are very reluctant to tell people how to feel, or what values they should have. I believe that an adequate analysis of rational choice cannot accept the evaluation of the consequences as given, and examine only the consistency of preferences. There is probably as much irrationality in our feelings, that is expressed in the way we evaluate the consequences, as there is in our choice of actions. An adequate normative analysis must deal with problems such as the legitimacy of regret in Allais' problem or the justifiability of the certainty effect described earlier in this lecture.

Discussions of the rationality of decisions in the context of utility theory are often misleading. Proponents and opponents of utility theory argue about the validity of certain axioms (e.g., substitutability, independence) where in fact the applicability of the axioms and not their adequacy is at stake. I do not see now the normative appeal of the axioms could be discussed without a reference to a specific interpretation. The axioms of utility theory can be regarded as maxims of rational choice only in conjunction with an intended interpretation, and the criteria for the selection of an interpretation are not part of utility theory. When Savage argues (convincingly, I believe) against Allais, he is arguing in effect for the monetary interpretation of the consequences as much as he argues for the independence axiom. Savage advises us, in effect, to disregard the element of regret and behave as if the effective consequences are limited to the monetary payoffs. In so doing, he is telling us how to feel and not how to choose. Personally, I find the argument compelling, but it is completely independent of utility theory. In this respect, Savage's analysis is therapeutic rather than prescriptive: It suggests a state of mind which leads to a particular resolution of the problem, but it contains no reason why that particular state of mind should be adopted.

In summary, the present lecture was devoted to a critique of expected utility theory from both descriptive and normative viewpoints. I have argued, on the basis of some recent evidence, that utility theory does not provide an adequate description of individual decision making under risk. In particular, the evidence indicates that both risk aversion and risk seeking are by-products of a more general phenomenon – the certainty effect – and that this effect cannot be accommodated by any (concave or convex) utility function. A careful examination of the question of rational choice under uncertainty reveals that it is not possible to evaluate the normative adequacy of the axioms of utility theory in the absence of a specific interpretation of the consequences. The criteria for the selection of such an interpretation, and for evaluating its appropriateness, however, are not part of expected utility theory. In this respect, utility theory provides only a partial analysis of the problem of rational choice. A comprehensive analysis of rational choice under risk should face the interpretation problem as well as the problem of the legitimacy of values. This analysis is likely to be explicative, or even therapeutic, rather than normative in nature. To the best of my knowledge, no such analysis is available at present.

NOTE

* Lecture given at the colloquium on 'Theories of Decision and their Semantics' in Salzburg, July 1974.

REFERENCES

Allais, M., 'Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine', *Econometrica*, 1953, **21**, 503–546.

Arrow, K. J., *Essays in the Theory of Risk-bearing*, Markham Publishing Company, Chicago, 1971.

Bernoulli, D., Specimen theoriae novae de mensura sortis, *Commentarii academiae scientiarum imperiales petropolitanae*, 1738, **5**, 175–192.

Davidson, D., Suppes, P., and Siegel, S., *Decision-making: an Experimental Approach*, Stanford University Press, Stanford, 1957.

Friedman, M., and Savage, L. J., The Utility Analysis of Choices Involving Risks, *Journal of Political Economy*, 1948, **56**, 279–304.

Mosteller, F. and Nogee, P., An Experimental Measurement of Utility, *Journal of Political Economy*, 1951, **59**, 371–404.

Pratt, J. W., Risk Aversion in the Small and in the Large, *Econometrica*, 1964, **32**, 122–136.

Raiffa, H., *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley, Reading, Mass., 1968.

Savage, L. J., *The Foundations of Statistics*, Wiley, New York, 1954.

Tversky, A., Additivity Analysis of Risky Choices, *Journal of Experimental Psychology*, 1967, **75**, 27–36.

Tversky, A., The Intransitivity of Preferences, *Psychological Review*, 1969, **76**, 31–48.

von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1944.