The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations

By Paul J. H. Schoemaker
Graduate School of Business
University of Chicago

The author would like to express his appreciation to David Brée, Victor Goldberg, Paul Kleindorfer, Roger Kormendi, Michael Rothschild, and the anonymous referees for their helpful comments on earlier drafts of this article. Additionally, his colleagues in the Center for Decision Research at the University of Chicago, especially Hillel Einhorn, Robin Hogarth and J. Edward Russo, are acknowledged for stimulating discussions on the present topic. Funding for this review was received from the Graduate School of Management in Delft and the Erasmus University of Rotterdam, both in the Netherlands. Their support is much appreciated.

I. Introduction

It is no exaggeration to consider expected utility theory the major paradigm in decision making since the Second World War. It has been used prescriptively in management science (especially decision analysis), predictively in finance and economics, descriptively by psychologists, and has played a central role in theories of measurable utility. The expected utility (EU) model has consequently been the focus of much theoretical and empirical research, including various interpretations and descriptive modifications as to its mathematical form. This paper reviews the major empirical studies bearing on the EU model. Although previous reviews of decision making have covered some of this research (e.g., Ward Edwards, 1961; Gordon Becker and Charles McClintock, 1967; Amnon Rapoport and Thomas Wallsten, 1972; Paul Slovic et al., 1977; Robert Libby and Peter Fishburn, 1977; Charles Vlek and Willem Wagenaar, 1979; and Hillel Einhorn and Robin Hogarth, 1981), few have attempted to organize the relevant evidence around the different purposes served by the EU model. Similarly, there has been no systematic examination of the way various descriptive extensions of expected utility theory relate to their progenitor, or of how the current normative variant differs from its historical roots.

In addressing these issues, the present paper first discusses various EU modifications. Special attention will be given to
the types of cardinal utility used in various models, as well as the manner in which probabilities are incorporated. Thereafter four conceptually different purposes of the EU model are identified, namely: descriptive, predictive, postdictive and prescriptive. The types of evidence relevant in testing the model's adequacy for each purpose are discussed. Building on these distinctions, the empirical evidence is then divided into four clusters centering around tests of the axioms, field research, information processing studies, and recent findings on context effects. Although the review covers considerable ground, its focus is on major studies; it does not provide a comprehensive discussion of all relevant research. Moreover, the focus is on individual decision making rather than the behavior of firms or markets. It shows that at the individual level most of the empirical evidence is difficult to reconcile with the principle of EU maximization. Whereas the simplicity of EU theory, especially its mathematical tractability, may make it a very attractive model for purposes of social aggregation, its structural validity at the individual level is questionable. As such, a separate section is devoted to important behavioral decision aspects that are currently ignored in EU theory. Finally, the discussion section synthesizes the divergent strands of research touched upon, with an eye to future roles of the EU model.

II. Expected Utility Variants

Expected utility models are concerned with choices among risky prospects whose outcomes may be either single or multidimensional. If we denote these various (say \( n \)) outcome vectors by \( \bar{x}_i \) and denote the \( n \) associated probabilities by \( p_i \) such that \( \sum_{i=1}^{n} p_i = 1 \), we then generally define an EU model as one which predicts or prescribes that people maximize \( \sum_{i=1}^{n} F(p_i) U(\bar{x}_i) \). The key characteristics of this general maximization model are: (1) a holistic evaluation of alternatives,\(^1\) (2) separable transformations on probabilities and outcomes, and (3) an expectation-type operation that combines probabilities and outcomes multiplicatively (after certain transformations).

Within this general EU model different variants exist depending on (1) how utility is measured, (2) what type of probability transformations \( F(\cdot) \) are allowed, and (3) how the outcomes \( \bar{x}_i \) are measured. In this section we examine some of the major variants, including their extra-mathematical interpretations. We start with some background information on EU theory as traditionally understood. Thereafter the neoclassical notion of cardinal utility is compared with the modern day one. Special attention is given to the ways in which EU theory can be cardinal. Finally the concept of probability is discussed, both in terms of its ontology and its treatment in various EU models. The section concludes with a summary table listing the major EU variants.

a. Background Information

The mathematical form of expected utility theory goes back as far as Gabriel Cramer (1728) and Daniel Bernoulli (1738), who sought to explain the so-called Petersburg paradox. The issue they addressed was why people would pay only a small dollar amount for a game of infinite mathematical expectation. This well-known game involves flipping a fair coin as many times as is necessary to produce "heads" for the first time. The payoff of this experiment depends on the number

\(^1\) A holistic model is one in which the attractiveness of an alternative is evaluated independently of the other alternatives in the choice set. In contrast, a non-holistic or decomposed model directly compares alternatives, e.g., one dimension at a time, without assigning a separate utility level to each.
of tosses required to get heads. Say this number is \( n \), the payoff will then be \( \$2^n \), which means the game has many possible outcomes, namely: \( \$2, 4, 8, \ldots, 2^n \), with probabilities \( \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \left(\frac{1}{2}\right)^n \), respectively. Interestingly, the expected monetary value (EV) of this gamble is infinite, since 
\[
\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n 2^n = \infty.
\]

To explain why most people value this infinite EV game below \( \$100 \), or even \( \$20 \), Bernoulli proposed that people maximize expected utility rather than expected monetary value. The utility function \( U(x) \) he proposed was logarithmic, exhibiting diminishing increases in utility for equal increments in wealth.\(^2\) Bernoulli then proceeded to show that for a logarithmic function the game’s expected utility, i.e., 
\[
\Sigma \left(\frac{1}{2}\right)^n \log_e(2^n),
\]

is indeed finite.\(^3\) However, he did not address the issue of how to measure utility, nor why his expectation principle would be rational. As such, Bernoulli’s theory is mostly a descriptive model, even though the expectation principle at the time may have enjoyed much face validity normatively. It was not until John von Neumann and Oskar Morgenstern (1944) that expected utility maximization was formally proved to be a rational decision criterion, i.e., derivable from several appealing axioms. In their own words they “practically defined numerical utility as being that thing for which a calculus of expectations is legitimate” (1944, p. 28). In this sense, von Neumann-Morgenstern (NM) utility theory is quite different from Bernoulli’s conceptualization. Moreover NM utility applies to any type of outcomes, money being a special case.

Specifically, von Neumann and Morgenstern proved that five basic axioms imply the existence of numerical utilities for outcomes whose expectations for lotteries preserve the preference order over lotteries: i.e., greater expected utility corresponds to higher preference. Their utility function is unique up to positive linear transformations, meaning that if the function \( U(x) \) represents a person’s risk preferences then so will \( U^*(x) \) if and only if \( U^*(x) = aU(x) + b \) for numbers \( a > 0 \) and \( b \). Jacob Marschak (1950) reformulated the NM axioms and proof, and proposed them as a definition of rational behavior under risk. In light of later discussions, the NM axioms are informally stated below.

1. Preferences for lotteries \( L_4 \) are complete and transitive. Completeness means that for any choice between lotteries \( L_1 \) and \( L_2 \) either \( L_1 \) is preferred to \( L_2 \) (denoted \( L_1 \succ L_2 \)), \( L_2 \succ L_1 \), or both are equally attractive. Transitivity implies that if \( L_1 \succ L_2 \) and \( L_2 \succ L_3 \) then \( L_1 \succ L_3 \) (where \( \succ \) denotes “at least as preferred as”).

2. If \( x_1 \prec x_2 \prec x_3 \), then there exists some probability \( p \) between zero and one such that the lottery \( p \frac{x_1}{1-p} \) is as attractive as receiving \( x_2 \) for certain.

3. If objects \( x_1 \) and \( x_2 \) (being either risky or riskless prospects) are equally attractive, then lottery \( 
\frac{p}{1-p} x_1 
\) and lottery \( 
\frac{p}{1-p} x_2 
\) will also be equally attractive (for any values of \( p \) and \( x_3 \)).
4. Consider the lotteries \( L_1 \) and \( L_2 \) which differ only in probability. If \( x_1 > x_2 \) then the first lottery \( (L_1) \) will be preferred over the second \( (L_2) \) if and only if \( p > q \).

5. A compound lottery (i.e., one whose outcomes are themselves lotteries) is equally attractive as the simple lottery that would result when multiplying probabilities through according to standard probability theory. For example, lottery

\[
\begin{array}{c}
L_1 \\
p \quad \frac{q}{1-p} \quad x_1 \\
\quad \frac{1-q}{1-p} \quad x_2 \\
\quad \frac{r}{1-r} \quad x_3 \\
\quad \frac{(1-p)r}{(1-p)(1-r)} \quad x_4
\end{array}
\]

should be as attractive as

\[
\begin{array}{c}
L_2 \\
pq \quad x_1 \\
\quad \frac{p(1-q)}{r} \quad x_2 \\
\quad \frac{(1-p)r}{(1-p)(1-r)} \quad x_3 \\
\quad \frac{(1-p)(1-r)}{(1-p)(1-r)} \quad x_4
\end{array}
\]

The above axioms are sufficient to guarantee that there exists a utility index such that the ordering of lotteries by their expected utilities fully coincides with the person’s actual preferences.\(^4\) Note that utility, in the NM context, is used to represent preferences whereas in neoclassical theory it determines (or precedes) preference. Since \( U(x) \) is unique up to positive linear transformation one is free to choose both the origin and the unit of measurement of the utility scale. For example, we may arbitrarily place the origin at \$10 (i.e., \( U(10) = 0 \)) and set \( U(10,000) \) equal to say \( 100 \) utiles. Given these two reference points, the utility index is constructed from such simple questions as: “What amount for certain is equally attractive as a 50–50 lottery offering \$10 or \$10,000?” Say the answer is \( \$x^* \), we then compute \( U(x^*) \) as being equal to \( .5U(10) + .5U(10,000) = 50 \) utiles. As long as the reference lottery contains amounts for which the utilities are known, new utility points can be obtained through which a utility function may then be interpolated.

An important concept in EU theory is that of risk aversion. If some gamble is less (or more) preferred than its expected monetary value for sure, the preference is said to be risk-averse (or risk-seeking). A concave utility function implies risk-averse preferences for lotteries within the range of concavity: i.e., their certainty equivalences will be less than their expected monetary values. Kenneth Arrow (1971) and John Pratt (1964) proposed as a local measure of risk-aversion for \( U(x) \) the negative ratio of the second to first derivative, i.e., \(-U''(x)/U'(x)\). This measure is invariant under linear transformation, and assumes a constant value for linear and exponential utility functions. As such it captures the important EU property that risk preferences derived from exponential (or linear) utility functions are not affected by changes in the person’s wealth position.\(^5\)

\(^4\) A highly readable proof of this important theorem was provided by William Baumol (1972, p. 548–51). Alternative sets of axioms, resulting in the same general theorem, have been presented by Israel N. Herstein and John Milnor (1953), Leonard Savage (1954), Duncan Luce and Howard Raiffa (1957), John Pratt, Raiffa and Robert Schlaifer (1964) and Peter Fishburn (1970).

\(^5\) Further discussions of expected utility theory, particularly from an applied perspective, can be found in Raiffa (1968) and in Chapter 4 of Ralph Keeney and Raiffa (1976). In subsequent chapters these authors extend the one-dimensional theory to \( n \)-dimensions, in which case the determination of \( U(x_1, x_2, \ldots, x_n) \) is considerably more complicated. Also, the concept of risk-aversion is less well-defined in the multiattribute case (Scott Richard, 1975).
b. Types of Cardinal Utility

As Peter Fishburn (1976) has noted, the concept of cardinal utility has psychological, empirical as well as measurement-theoretic aspects which together with such related terminology as “measurable,” “additive,” “determinate,” “intensive,” and “linear” utility has given rise to considerable confusion as to its precise meaning. The term “cardinal utility” goes back to John R. Hicks and R. G. D. Allen (1934) who argued that only ordinal preference was needed in economic theory, thereby dispensing with neoclassical utility (Vivian Walsh, 1970). Cardinal utility in the neoclassical context refers to strength of preference, i.e., to statements about intensity as well as direction of preference. From a measurement-theoretic viewpoint, cardinal utility has a rather different meaning, referring to the allowable transformations of the underlying measurement scale. If the scale is unique up to at least linear transformation, it constitutes cardinal or so-called strong measurement (S. S. Stevens, 1946). Common examples are temperature and weight measures which constitute interval and ratio scales respectively. From a measurement perspective NM utility theory is cardinal in that its utility scale has interval properties. However, from a preference perspective, NM utility theory is ordinal in that it provides no more than ordinal rankings of lotteries.

The cardinal nature of NM theory must thus be interpreted carefully. Even though NM utility functions are interval scales, implying that the ratios of utility differences are invariant under linear transformations, it does not follow that if \( x_1 > x_2 > x_3 > x_4 \) and \( u(x_1) - u(x_2) > u(x_3) - u(x_4) \), the change from \( x_2 \) to \( x_1 \) would be more preferred than the change from \( x_4 \) to \( x_3 \) (Duncan Luce and Raiffa, 1957, p. 32). Thus NM utility should not be interpreted as measuring strength of preference under certainty, being quite different in this regard from neoclassical cardinal utility (George Stigler, 1950). One reason is that preferences among lotteries are determined by at least two separate factors; namely (1) strength of preference for the consequences under certainty, and (2) attitude toward risk. The NM utility function is a compound mixture of these two, without direct resort to interval comparisons or strength of preference measures. As a preference theory, it is wholly ordinal. Nevertheless it implicitly assumes that a neoclassical type of utility exists, otherwise it would not be possible psychologically to determine the certainty equivalence of a lottery. An interesting analysis as to the connection between ordinal and cardinal utility was offered by Eugene Fama (1972). Since some economists consider intensity of preference meaningless (Charles Plott, 1976, p. 541), putatively because it cannot be measured from revealed preferences, it merits closer examination.

One approach is to view strength of preference as an intuitive psychological primitive. For instance, most people would consider it meaningful to say that the increase in pleasure due to adding milk to one’s coffee is of a lower magnitude than the pleasure increment associated with a sizable salary raise. Similarly, someone might note that the last hour on some trip was more tiring than the first. Indeed, in psychological scaling experiments subjects routinely make interval comparisons involving such quantities as loudness, weight, temperature, and brightness (Stevens, 1957). Usually, subjects’ perceptions of the interval differ-

\[\text{Consequently, the notion of marginal utility has a rather different meaning in NM theory as well. In classical economics marginal utility refers to pleasure increments under certainty. In NM theory it refers to “the marginal rate of substitution between } x \text{ and the probability of winning the prespecified prize of the standard lottery ticket.” (Baumol, 1972, p. 548).}\]
ences correspond very closely (in the curve fitting sense) to the true underlying scale, after appropriate logarithmic or power transformations. The latter are necessary as the human response system tends toward relative rather than absolute judgments (Gustav Fechner, 1860). It is thus a small leap to presume a similar judgment capability for strength of preference, even though objective verification is not yet possible.

Gerard Debreu (1959), D. Scott and Patrick Suppes (1958), Ragnar Frisch (1964), Frantz Alt (1971), and others, proposed various axiomatizations of such a strength of preference measures under certainty, which we shall denote by \( v(x) \). Its essential property is that the function provides ordinal preference as well as an ordering on differences (under certainty). Thus, \( v(x_1) > v(x_2) \) implies that \( x_1 \) is preferred to \( x_2 \), and \( v(x_1) - v(x_2) > v(x_3) - v(x_4) \) implies that the value difference between \( x_1 \) and \( x_2 \) is greater than that between \( x_3 \) and \( x_4 \) (where \( x_3 \) is preferred to \( x_4 \)).

David Krantz, et al., (1971, pp. 145–50) review the formal properties of this so-called positive-difference structure. Operationally, it may rely on the so-called midpoint scaling technique, which requires respondents to split intervals into equally valued increments (Warren Torgerson, 1958).

A different measurement approach is to infer \( v(x) \) from revealed preferences, provided certain conditions hold. By introducing a second attribute, say \( y \), it may be asked how much of \( y_0 \) (the initial \( y \) endowment) the respondent would give up to go from \( x_2 \) to \( x_1 \). From such willingness-to-pay questions a multidimensional interval-scaled function \( W(x, y, z) \) can be constructed via conjoint measurement (Luce and John Tukey, 1964), with \( z \) denoting the set of other relevant attributes. If \( W(x, y, z) \) is separable in \( x \), meaning it can be written as \( f(v(x), w(y,z)) \), and if \( \delta W/\delta x \) is independent of \( y \) and \( z \), then \( v(x) \) may be considered an intrinsic preference measure for \( x \). These conditions are met, for example, if \( W(x, y, z) \) is additive, i.e., expressible in the form \( v(x) + w_1(y) + w_2(z) \). However, a direct empirical test that \( \delta W/\delta x \) is independent of \( y \) and \( z \) may require another (i.e., non-tradeoff) measure of strength of preference (David Bell and Raiffa, 1979). If so, we are back to our first approach, leaving introspection as the only likely way out of this vicious circle (Fishburn, 1970, p. 82). Various other operational measures exist for intensity of preference, which were recently examined by John Hauser and Steven Shugan (1980). Their study consists of theoretical and empirical analyses, with a focus on marketing applications.

There are several advantages in distinguishing cardinal utility measures constructed under certainty, denoted \( v(x) \), from those constructed under risk, denoted \( u(x) \). First, it emphasizes that there exist different types of cardinal utility, even within each category, which only have to be related monotonically. (See Amos Tversky, 1967, for empirical examples.) Second, by examining \( u(x) = f(v(x)) \), an Arrow-Pratt type measure of intrinsic risk aversion may be defined and empirically measured, namely \( -f''(v(x)) / f'(v(x)) \) (Bell and Raiffa, 1979). Third, the construction of \( u(x) \) may be simplified by first examining the nature of \( v(x) \), especially in the case of multiattribute utility. For example, Detlof von Winterfeldt (1979) proved that \( u(x_1, \ldots, x_n) \) must be either linearly, logarithmically or exponentially related to \( v(x_1, \ldots, x_n) \) in case the former is additive and the latter multiplicative (see also: James Dyer and Rakesh Sarin, 1979a). Similar relationships between riskless and risky cardinal utility were recently examined by Bell and Raiffa (1979) and Dyer and Sarin (1982) for one-dimensional cases (see also: Sarin, 1982). Dyer and Sarin (1982) proposed a fourth reason
for separating $v(x)$ and $u(x)$: namely, group decision making. In an organizational context it may be desirable only to have members’ inputs regarding their $v(x)$ functions, but not their risk attitudes (which might be centrally determined). Finally, cardinal utility under certainty may be useful for welfare theory (Dyer and Sarin, 1979b), although it seems to suffer as well from impossibility theorems (T. Schwartz, 1970).

The distinction between $v(x)$ and $u(x)$, denoting cardinal preference scalings under certainty and risk, respectively, is often overlooked or has been a source of considerable confusion, even among experts. During the beginning of the 1950s, The Economic Journal and Econometrica published a variety of articles debating the cardinality of NM utility (e.g., Herman Wold, 1952 and Armen Alchian, 1953). Amid considerable confusion, lucid analyses were offered by Robert Strotz (1953), Daniel Ellsberg (1954) and John S. Chipman (1960). Ellsberg compared how such classical utilitarians as William Stanley Jevons, Carl Menger, Leon Walras or Alfred Marshall might have predicted choice under risk with the approach taken by NM. The difference lies in the way the utility function is constructed: namely, under certainty or risk. Which expectation model will predict better is an empirical question. An important difference, however, is that the $E[v(\bar{x})]$ model has no normative justification other than its face validity, whereas the $E[u(\bar{x})]$ model derives from a set of appealing decision axioms. It is my interpretation that Bernoulli proposed the $v(x)$ type expectation model, although he never explicitly addressed the measurement question.

The above discussion on cardinal utility was provided because, even today, the distinction between $v(x)$ and $u(x)$ is often unrecognized. Textbooks in economics and management science occasionally discuss the NM function as if it only measured intrinsic pleasure under conditions of certainty. For example, a concave $u(x)$ might erroneously be interpreted as implying that equal increments in money (under certainty) contribute to utility at a decreasing rate. Of course, $v(x)$ is confused here with $u(x)$. Apart from textbook confusions, the above distinction is occasionally not recognized in research designs. For example, in a study of simulated real-life decisions, such as whether or not to enroll in a Ph.D. program, Thomas Bonoma and Barry R. Schlenker (1978) considered subjects suboptimal if they did not maximize $\Sigma p_i v(x_i)$; what these authors should have tested, in making such normative evaluations of risky choices, is whether $\Sigma p_i u(x_i)$ was maximized.

To summarize, $v(\bar{x})$ was defined as an interval-scaled utility measure constructed under conditions of certainty, similar to neoclassical utility except that no ratio properties are presumed. In contrast, $u(\bar{x})$ is a cardinal NM utility measure derived from preferences among lotteries. These two utility functions only need to be monotone transforms of each other. Thus when presented with different commodity bundles $\bar{x}_i$ under certainty, $v(\bar{x})$ and $u(\bar{x})$ should yield the same ordering for a person satisfying the axioms underlying either construct. However, when presented with risky prospects $\bar{x}_i$, the formally correct ordering is determined from $E[u(\bar{x}_i)]$, which will generally differ with that obtained from $E[v(\bar{x}_i)]$ unless $v(x)$ is a linear transform of $u(x)$. Finally to obtain an interval ranking of the risky prospects $\bar{x}_i$, the $E[u(\bar{x}_i)]$ could be inverted into their certainty equivalents $CE_i$ which might then be interval ranked by computing $v(CE_i)$.

c. The Concept of Probability

Other potentially confusing aspects of the expected utility model concern the treatment of probabilities. In the NM axiom system probability is considered a
primitive whose values are objectively given. Empirically, however, the notion of probability is problematic, both philosophically as well as practically. To illustrate this let us briefly identify four major schools of probability and their limitations. The first is the classical view of Pierre Laplace (1812), who defined probability as the number of elementary outcomes favorable to some event divided by the total number of possible elementary outcomes. Since these elementary outcomes must all be equally likely (i.e., have the same probability), Laplace's definition suffers from circularity. Moreover, this view cannot easily handle infinite outcome spaces, and is practically limited to well-structured uncertainties.

Jacques Bernoulli (1713), a relative of Daniel, had earlier evaded this circular definition by distinguishing the concept from its measurement. He defined probability as a "degree of confidence" which for a given event may vary from person to person. Nevertheless, he considered the art of guessing (Ars Conjectandi) to consist of precise estimation of unknown probabilities, for instance by studying objective frequencies. This frequency approach was later placed on an axiomatic footing by John Venn (1866), Hans Reichenbach (1935), and Richard Von Mises (1957, 1964) who defined probability as the limiting value of the percentage of favorable outcomes in an infinite sequence of independent replications. The limitations of this view are at least three-fold. First, probability is never exact numerically, being at best a large sample estimate. Second, it is often unclear what sample space to use. For instance, when taking a plane trip is the objective probability of crashing determined by all previous flights, or only those by that carrier, with that type of airplane, in that season, etc.? Third, the notion of exact replication is problematic. If a coin toss is replicated perfectly, it should always yield the same outcome. This raises questions about the locus of uncertainty; i.e., whether it is internal or out in the world. The answer depends on one's world-view; for some people determinism may rule out true randomness (its only source being imperfect knowledge), whereas others might argue in favor of irreducible uncertainty: e.g., Heisenberg's uncertainty principle in physics (R. B. Lindsay, 1968).

A third attempt to define probability objectively is the so-called logical school of John Maynard Keynes (1921) and Harold Jeffreys (1948). These authors argued that a given set of evidence bears a logical, objective relationship to the truth of some hypothesis (e.g., someone being guilty), even when the evidence is inconclusive. Probability measures the strength of this connection as assessed by a rational person. Since all three of the above views have appealing aspects various attempts have been made to reconcile them. Rudolf Carnap (1962, 1971) developed a formal theory of a coherent learning system (along Bayesian lines), which merges objective and subjective views. Glenn Shafer (1976) on the other hand, proposed a reconciliation by formally distinguishing different types of probability, emphasizing the essential difference between aleatory probability and degree of belief. This latter epistemic concept is fundamental to the subjectivists, the fourth school to be mentioned.

The subjective or personal school of probability was primarily developed by Frank Ramsey (1931), Bruno de Finetti (1937, 1974), Leonard Savage (1954), and Pratt, Raiffa and Schlaifer (1964). In their view, probabilities are degrees of beliefs, applicable to both repetitive and unique events such as a third world war. For a given set of hypotheses, any assignment of subjective probabilities is permissible in principle, provided some consistency requirements are met. In contrast to other schools, these conditions are viewed as suf-
ficient (as well as necessary), without additional restrictions being imposed for logical or empirical reasons. The main consistency axiom of subjective probability theory is coherence (de Finetti, 1937). Informally, it requires that for a given belief system, it should not be possible for a clever bookmaker to lay multiple fair bets so that the bookmaker wins under all possible outcomes. This axiom (together with some others) implies that the probabilities of elementary events sum to one, and that conjunctive and disjunctive events follow the product and addition rules respectively. As such, subjective probabilities are mathematically indistinguishable from other types of probability. The subjective school developed measurement procedures for the simultaneous estimation of utility and probability as based on revealed preferences (Donald Davidson and Suppes, 1956).

The important point is that probability is not a simple construct (Henry Kyburg and Howard Smokler, 1964). Its measurement is obviously difficult in real-world settings, but may even be so in simple games of chance (Davidson, Suppes and Sidney Siegel, 1957). To distinguish subjective from objective probability, it shall be denoted by \( f(p) \). The \( f(\cdot) \) transformation signals that the probabilities to be used in EU model may differ from the stated or objective ones assumed by the researcher. However, not all \( f(p_i) \) transformations satisfying the properties of probability (such as \( \Sigma f(p_i) = 1 \)) should be viewed as degrees of belief. In the literature, \( f(p_i) \) transformations have been proposed to reflect risk-taking attitudes (Jagdish Handa, 1977), to explore symmetry between the probability and outcome component in expectation models (Hans Schneeweiss, 1974), to reflect probability and/or variance preference (Edwards, 1954a and 1954b), or simply to fit data under the assumption that preferences are non-linear in probability (John Quiggin, 1980). Although these various models are commonly referred to as Subjective Expected Utility theory (SEU), the \( f(p_i) \) transformation may not strictly be degree of belief measures.

Apart from transformations that do preserve the properties of mathematical probability, various theories exist where this requirement is relaxed. In Table 1 such probability transformations are denoted by \( w(p_i) \), and will be called decision weights. To quote Daniel Kahneman and Tversky (1979, p. 280), “decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree of belief.” In prospect theory (Kahneman and Tversky, 1979) decision weights are aimed at reflecting the impact of events on the overall attractiveness of gambles. As such they are monotonic with probability, but not necessarily linear.

To summarize, there are several ways that utility and probability have been treated in EU models. Systematic combining of the different transformations discussed yields nine EU variants, which are shown in Table 1 together with their names and main originators. Note that this table focuses on the allowable probability and outcome transformations of the various models. However, there are other differences as well. For example, in prospect theory the outcomes \( x_i \) are defined on changes in financial position rather than on final asset position. Moreover, in descriptive models the outcome space may include such dimensions as regret, justifiability of one’s choice, etc. Most of the models listed were advanced as descriptive ones, with the exception of von Neumann-Morgenstern and Savage. It is

---

7 The subjective expected utility (SEU) model is also encountered outside the context of monetary bets. Victor Vroom’s (1964) expectancy theory of work motivation shows formal parallels to EU theory, as do theories in learning, attitude formation, and personality development (Edward Lawler, 1973).
these normative variants, especially NM theory, we henceforth refer to when speaking of EU theory:

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nine Variants of the Expected Utility Model</td>
</tr>
<tr>
<td>1. $\Sigma p_i x_i$</td>
</tr>
<tr>
<td>2. $\Sigma p_i v(x_i)$</td>
</tr>
<tr>
<td>3. $\Sigma p_i u(x_i)$</td>
</tr>
<tr>
<td>4. $\Sigma f(p_i) x_i$</td>
</tr>
<tr>
<td>5. $\Sigma f(p_i) v(x_i)$</td>
</tr>
<tr>
<td>6. $\Sigma f(p_i) u(x_i)$</td>
</tr>
<tr>
<td>7. $\Sigma w(p_i) x_i$</td>
</tr>
<tr>
<td>8. $\Sigma w(p_i) v(x_i)$</td>
</tr>
<tr>
<td>9. $\Sigma w(p_i) u(x_i)$</td>
</tr>
</tbody>
</table>

Note: $v(x)$ denotes an interval scaled utility measure constructed under certainty; $u(x)$ denotes one constructed via lotteries.

III. Purposes of the EU Model

Generally, it would be inappropriate to assess the acceptability of an optimality model without an explicit prior statement of its purposes. If models are defined as simplified representations of reality, they should not and cannot always be true. Models, by their nature, balance costs such as complexity, prediction error, domain specificity, with various benefits such as simplicity, prediction power, generality, etc. Consequently, we should evaluate a model according to its stated objectives. We shall distinguish four essentially different purposes that the EU model can serve.

First, EU theory may be used descriptively to model the decision processes underlying risky choice. Descriptive models are concerned with, and tested by more than prediction alone. Evidence concerning the validity of the underlying axioms would be relevant, as would the manner in which information is processed. The latter may include: (1) how stimuli are attended to, encoded and stored; (2) how information is searched for and retrieved from memory, (3) how stimuli are aggregated or decomposed (i.e., comprehended and integrated), and (4) how value conflicts are resolved when choice is exercised (James Bettman, 1979).

A second usage, dominant in economics and finance is to view the EU model as predictive or positivistic. Realism of its axioms and postulated computational mechanism are not important within the positivistic view. What matters is whether the model offers higher predictive accuracy than competing models of similar complexity. According to Milton Friedman (1953) and Fritz Machlup (1967), two leading exponents of the positivistic view, (economic) theories should be tested on their predictive ability rather than the descriptive validity of their assumptions. Hence direct violations of EU axioms are not particularly disturbing. What counts is whether the theory, in its capacity of an “as if” model, predicts behavior not
used in the construction of the model. In that case, empirical findings of field studies and realistic laboratory studies are especially relevant. However, in such studies the evidence counter to EU theory can often be refuted because it is indirect.

For instance, in field studies assumptions need to be made about the values of various parameters (e.g., costs, tax rates, probabilities, etc.) as well as the form of $U(x)$. If the data are inconsistent with certain of these prior assumptions, another set of parameters can usually be found, ex post facto, that fit the data better. To illustrate, when Peter Pashigian, Lawrence Schkade and George Menefee (1966) concluded that consumers did not act in accordance with EU theory because they bought expensive low-deductible policies for collision in automobile insurance, John Gould (1969) countered that the data were in fact consistent with certain types of sufficiently risk-averse exponential utility functions. The presumed incompatibility only applied to quadratic and logarithmic utility functions. The question therefore arises, within the positivistic view, to what extent the EU hypothesis is falsifiable. The most irrefutable evidence, which directly concerns the axioms, is discounted in positive economics. On the other hand, the evidence that is admissible (e.g., actual decision making in the real world) often suffers from lack of controls, multiple interpretations, and measurement problems regarding key constructs (e.g., probabilities). Some examples are provided later in the context of the capital asset pricing model in finance and studies on the effect of pension plans on private savings.

Nevertheless, the distinguishing feature of a positive theory is that it yields hypotheses that are falsifiable in principle: i.e., they meet Karl Popper’s (1968) so-called demarcation criterion. The latter requires that a theory specifies clearly and a priori the conceivable empirical results which support it and those which refute it (without the latter set being empty). To distinguish between those researchers who allow falsification of the EU model in principle, and those who regard the optimality of economic behavior as an essentially unfalsifiable meta-postulate, we might distinguish a third purpose of EU models, which I shall refer to as postdictive. The essential premise of the postdictive EU view is that all observed human behavior is optimal (in the EU sense), provided it is modeled in the appropriate manner. Seeming suboptimalities are explained, ex post facto, by introducing new considerations (e.g., costs, dimensions, constraints, etc.) that account for the anomalies, so as to make them optimal. From this perspective, satisficing (Herbert Simon, 1955) is just a more general type of optimizing, including such factors as the cost of information, decision time, constraints, and cognitive effort. It is the latter degrees of freedom, however, that may make the postdictive approach tautological: i.e., non-empirical and non-falsifiable (for a different view see Lawrence Boland, 1981).

Many economists have acknowledged the postdictive perspective in economics, some with sorrow and some with pride. For instance, Tibor Scitovsky (1976) remarked that the concept of utility maximization “set back by generations all scientific inquiry into consumer behavior, for it seemed to rule out—as a logical impossibility—any conflict between what man chooses to get and what will best satisfy him” (p. 4). Gary Becker (1976) on the other hand views this as a strength; he acknowledges the potential for tautology or circularity in economics, especially if unobservable transaction costs are permitted ex post: “of course, postulating the existence of costs closes or ‘completes’ the economic approach in the same, almost tautological, way that postulating the existence of (sometimes unobserved) uses of energy completes the energy system and
preserves the law of the conservation of energy . . . the critical question is whether a system is completed in a useful way" (p. 7), i.e., whether it yields "a bundle of empty tautologies" or provides the basis for predicting behavioral responses to various changes.

The challenge of the postdictive view is to look at the available data in a coherent way that highlights the optimality of human behavior (e.g., Becker, 1976). Whether such an exercise is trivial or a masterly craft depends on the degrees of freedom one allows. The possible respecifications of a particular model are of course not arbitrary and must gain the approval of fellow practitioners. Furthermore, most model respecifications imply testable predictions. However, if these testable predictions were also not to hold, additional respecifications would be sought until all relevant past (i.e., observed) behavior would be accounted for as indeed being optimal. It is in this sense that the perspective is ex post. A major limitation of the postdictive view is that ex post empirical models may have limited refutation power regarding the corresponding theoretical ex ante model. The Sharpe-Lintner capital asset pricing model in finance (see Fama, 1976) is a case in point. Its main hypothesis of a linear relationship between a security's return and the market's return has evoked numerous empirical studies. However, this hypothesis presumes that the market portfolio has minimum variance for its level of expected return (i.e., is mean-variance efficient). Since the market portfolio includes all assets (financial and otherwise) it is in reality unknowable, making empirical tests very difficult if not impossible (Richard Roll, 1977).

This is not to say, however, that closed systems are without merit. The postdictive view resembles the natural sciences in their ex post search for optimality principles, some of which have proved most valuable. For instance, when Pierre Fermat suggested (c. 1650) that light travels the path of the shortest distance in time, he not only derived Willebrord Snel's refraction law in optics (which was his original intent), but also suggested a host of new hypotheses quite unrelated to the refraction law. Many of these hypotheses were subsequently verified empirically, dealing with relative velocities of light in different media and the behavior of light in convergent lenses. Two centuries later, William Hamilton generalized and elevated the least effort principle to a cornerstone theorem in theoretical physics. Similarly, in biology optimality principles abound, ranging from homeostasis (i.e., minimization of the divergence between actual and desired states) to optimal adaptation (Robert Rosen, 1967; M. Cody, 1974 or J. Maynard Smith, 1978). Elsewhere I examined optimality principles in more detail (1982), and concluded that their main advantages are: (1) elegance, (2) parsimonious summary of empirical knowledge, and (3) high metaphoric value in generating new hypotheses. The disadvantages, on the other hand, are that the optimality approach (1) encourages search for confirming rather than disconfirming evidence, (2) often fails to acknowledge its ex post nature, and (3) may be more reflective of analytical than empirical truths. Thus, when sociobiology (Edward Wilson, 1975) broadly claims to have established that the existing structures in human and animal societies benefit the stronger genes, Richard Lewontin (1979) justifiably highlights the difficulty of defining such key terms as "favorable traits," "ecological niche" and "strong genes," independently of knowing evolution's outcomes. Similarly, when economics expands its claim of optimal choice behavior into new and broad domains of human activity (e.g., Gary Becker, 1976), or even animal behavior, the tautological nature of these models (in the sense of ad hoc,
ex post rationalization) may well increase accordingly.\textsuperscript{8} It is beyond our scope to examine further the merits of positivistic and postdictive perspectives as they entail complex epistemological issues. Penetrating analyses on this topic were recently offered in books by Alexander Rosenberg (1975) and Mark Blaug (1980).

Finally, there exists a fourth perspective according to which EU theory is a \textit{prescriptive} or \textit{normative} model. Decision analysts and management scientists (implicitly) assume that human behavior is generally suboptimal. Their goal is to improve decisions (prescriptive), for instance by using EU theory (normative). If so, the theory serves to advise which alternative(s) to select in complex decision situations on the basis of the decision maker's basic tastes and preferences. As outlined earlier these basic risk preferences are captured through a NM utility function. The complex options are then rank-ordered on the basis of expected utility.

Each of the four EU purposes discussed above has its constituents, who would interpret empirical evidence on the EU model differently. For instance, evidence on axioms is relevant to the descriptive and prescriptive usages whereas predictions, preferably about real-world behavior, matter to the predictive or postdictive views. A complicating feature of the EU model as employed in economics and finance (e.g., Peter Diamond and Rothschild, 1978) is that the assumption of EU maximization may have the following three important qualifiers. First, the theory is usually restricted to decisions that are important economically. Evidence on hypothetical decisions might thus be dismissed on grounds of having no significant economic consequence to the decision maker. Second, since the concern in economics is with market rather than individual behavior, only those individuals trading at the margin need to be EU maximizers. For instance, in busy traffic only a small portion of highway drivers need to change from slower to faster lanes (a type of arbitrage) for the lanes to move at the same speed eventually. Thus it often requires only a few rational persons for the market as a whole to behave rationally. Third, in economics, EU maximization is mostly assumed in \textit{competitive} environments, where feedback enables people, over time, to improve their behavior. Hence, some history of learning and struggle for economic survival are often assumed for the EU model as used in finance and economics (Alchian, 1950; Sidney Winter, 1964 and 1971).

IV. Empirical Evidence

In recognition of the various purposes served by EU models, the present overview of empirical evidence is organized around four different approaches. Although some research studies could properly be discussed under several of these, they are mostly treated under just one category.

\textit{a. Tests of EU Axioms}

The first EU axiom has two components: (1) for any choice people have a definite preference (including indifference), and (2) preferences are transitive. In one of the first empirical tests of the EU model (1951) Frederich Mosteller and Philip Nogee found, however, that subjects on repeated measures of preference would not always give the same answers. They subsequently conducted their research assuming stochastic rather than deterministic

\textsuperscript{8} This is not to say that economic animal experiments advance non-falsifiable hypotheses. For instance, Raymond Battalio, et al. (1981) found in their experiments that pigeons engaged in insufficient substitution when compared to the Slutsky-Hicks theory. However, in some respecified model the observed behavior might well be "optimal." My concern is with the meta-postulate that all animal behavior is in some sense optimal, as this may lead to empirically vacuous theories.
preferences. Concerning the transitivity axiom, Tversky (1969) examined some conditions under which it might be violated. The axiom has a deterministic and stochastic form: in the former it states that $A \succ B$ and $B \succ C$ implies $A \succ C$. In its (weak) stochastic form it asserts that if the probability that $A \succ B$, denoted $P(A \succ B)$, is at least $\frac{1}{2}$, and $P(B \succ C) \geq \frac{1}{2}$, then $P(A \succ C) \geq \frac{1}{2}$. Violations of stochastic transitivity cannot be attributed to random error.

In two separate experiments, one dealing with gambles, the other with college applicant decisions, Tversky (1969) showed systematic and predictable violations of weak stochastic transitivity. The violations are likely to occur if subjects use evaluation strategies involving comparisons within dimensions: e.g., first comparing price, then quality, then size, etc. The strategy Tversky examined is the so-called additive difference model in which an alternative $\overline{x} = (x_1, \ldots, x_n)$ is preferred to $\overline{y} = (y_1, \ldots, y_n)$ if and only if $\sum_{i=1}^{n} \phi_i[u_i(x_i) - u_i(y_i)] \geq 0$. The $u_i$ functions measure the subjective values of the various attributes, and the $\phi_i$ are increasing continuous functions determining the contribution of each subjective difference within dimension $i$ to the overall evaluation of the alternatives. Tversky proved analytically that for $n \geq 3$, transitive choices are guaranteed under the additive difference rule if and only if the $\phi_i$ are linear, which is a rather severe restriction.

The second EU axiom (combined with that of transitivity) implies that a lottery offering outcomes A and B should have an attractiveness level intermediate to those of A and B. This in-betweenness property was experimentally tested by Clyde Coombs (1975). Subjects were asked to rank three gambles A, B, and C in order of attractiveness, where C was a probability mixture of A and B. For example, if A offers a 50–50 chance at $3 or $0, and gamble B offers a 50–50 chance at $5 or $0, then a 40–60 mixture of A and B, called gamble C, would offer outcomes of $5, $3 and $0 with probabilities .3, .2 and .5 respectively. According to the EU axioms, gamble C should be in-between A and B in terms of attractiveness. In evaluating these three gambles, subjects can give six possible rankings, which Coombs reduced to three basic classes, namely monotone orderings (ACB or BCA), folded orderings (CAB or CBA), and inverted orderings (ABC or BAC). Only the first of these, the monotone ranking, is consistent with EU. Of the 520 rank-orderings Coombs examined, 54 percent were monotone, 27 percent folded, and 19 percent inverted, suggesting that nearly half the subjects violated the in-betweenness axiom. Similar violations had been observed earlier by Becker, Morris de Groot, and Marschak (1963).

The third axiom, which assumes invariance of preference between certainty and risk when other things are equal, was examined by Kahneman and Tversky (1979) as a generalization of Maurice Allais’ (1953) well-known paradox. They observed a so-called certainty effect according to which outcomes obtained with certainty loom disproportionately larger than those which are uncertain. As an example, consider the following two-choice situations:

**Situation A:**

(1a) a certain loss of $45.
(2a) a .5 chance of losing $100 and a .5 chance of losing $0.

**Situation B:**

(1b) a .10 chance of losing $45 and a .9 chance of losing $0.
(2b) a .05 chance of losing $100 and a .95 chance of losing $0.

Most subjects preferred (2a) to (1a) and (1b) to (2b) which violates EU since the former implies that $U(-45) < .5U(-100)$
+ .5U(0), whereas the latter preference implies the reverse inequality. To see why this preference pattern violates the third EU axiom, note the following relationship between situations A and B:

\[
(1b) = \frac{.1}{.9} \left(1a\right) \quad \quad \quad (2b) = \frac{.1}{.9} \left(2a\right)
\]

In prospect theory, Kahneman and Tversky (1979) provide several additional examples of this certainty effect, which they explain as the result of probability distortions.

A related experiment on the effect of certainty vs. uncertainty was performed by Ellsberg (1961); however, his focus was on the probability dimension. He showed that people dislike ambiguity as to the exact level of the probability of winning (p_i). The essential finding was that subjects look at more than the expected value of such a probability, when given some prior distribution as to its levels. This sensitivity to other moments is contrary to EU theory for which only \(E(\hat{p}_i)\) matters. The Allais and Ellsberg paradoxes can both be shown to entail a violation of an alternative type of EU axiom, called Savage's independence or sure-thing principle (1954). This axiom holds that if two lotteries have an identical probability and payoff branch, the levels of this payoff and probability should not affect people's choice between the lotteries.

The above violations particularly raise questions about the fifth EU axiom, i.e., the assessment of compound lotteries. Direct evidence of its violation was offered by Maya Bar-Hillel (1973) who found that people tend to overestimate the probability of conjunctive events and underestimate the probability of disjunctive events. In her experiment subjects were presented with three events on which they could bet:

1. A simple event: drawing a red marble from a bag containing 50 percent red and 50 percent white marbles.
2. A conjunctive event: drawing a red marble 7 times in succession, with replacement, from a bag containing 90 percent red marbles and 10 percent white marbles.
3. A disjunctive event: drawing a red marble at least once in 7 successive tries, with replacement, from a bag containing 10 percent red marbles and 90 percent white marbles.

The majority preferred bet 2 to bet 1. Subjects also preferred bet 1 to bet 3. The probabilities of the three events, however, are .50, .48 and .52 respectively. These biases can be explained as effects of anchoring: i.e., the stated probability of the elementary event (.1 in the disjunctive case) provides a natural starting point from which insufficient adjustment is made to arrive at the final answer.

A sixth postulate, not formally part of NM utility theory, is the traditional risk-aversion assumption in economics and finance, particularly for losses (Friedman and Savage, 1948; Arrow, 1971; Isaac Ehrlich and Becker, 1972; J. Marshall, 1974; Martin Bailey et al., 1980). Recent laboratory studies, however, seriously question this pervasive assumption (also see, Harry Markowitz, 1952). For instance, John Hershey and Schoemaker (1980) found that less than 40 percent of their subjects would pay $100 to protect against a .01 chance of losing $10,000. Similar refusals to accept actuarially fair insurance were uncovered by Paul Slovic et al. (1977), Kahneman and Tversky (1979), Schoemaker and Howard Kunreuther (1979), and Fishburn and Gary Kocheberger (1979). Dan Laughhunn et al. (1981) found

---

9 A normative reply to the Ellsberg paradox can be found in Raiffa (1961), Harry Roberts (1963), and Jacques Drèze (1974). Other discussions of the Ellsberg experiment and related paradoxes (e.g., Allais) are contained in William Fellner (1961), Selwyn Becker and Fred Brownson (1964), Karl Borch (1968), Kenneth MacCrimmon (1968), Frank Yates and Lisa Zukowski (1976), MacCrimmon and Stig Larsson (1979), and Soo-Hong Chew and MacCrimmon (1979b).
executives to be risk-seeking when presented with gambles having outcomes below target return levels (e.g., breakeven returns on investments). However, when the gambles involved ruinous losses, the executives were often risk-averse.

b. Field Studies

From a positivistic viewpoint the above evidence may be of limited interest. For instance, economists might argue that although subjects may behave non-optimally in hypothetical and unrepresentative experiments, in the real world, where decisions do matter, people will in fact maximize expected utility.

Real-world data on insurance against flood and earthquake insurance, however, suggest a different conclusion (Dan Anderson, 1974). In spite of Federal subsidies of up to 90 percent for flood insurance (this past decade), Kunreuther et al. (1978) found that the majority of eligible homeowners in flood plains were uninsured. One obvious reason could be that people’s perceptions of the probability and magnitude of loss, or their perceptions as to the cost of insurance, differed from the actuarial figures used by the government. To test whether homeowners were subjectively rational, Kunreuther et al. interviewed 2,000 homeowners in flood plains and 1,000 homeowners in earthquake areas to obtain subjective estimates of these figures. Many respondents, over one-half, were ill-informed on the availability of insurance against these hazards. Of those who were aware, however, many acted contrary to subjective EU maximization (around 40 percent for flood and 30 percent for earthquake insurance), assuming a marginal tax rate of 30 percent and plausible degrees of risk aversion. The study also controlled for people expecting post-disaster relief aid from federal, state or local agencies. Results found by Kunreuther et al (1978) thus seriously question people’s ability to process information on low probability, high loss events. (See also Kunreuther, 1976). A similar conclusion emerges from the low demand for crime insurance (Federal Insurance Administration, 1974), and people’s general reluctance to wear seat belts in automobiles (L. Robertson, 1974).10

Although the above results imply risk-seeking for losses (within the EU model), in other domains of insurance people appear highly risk-averse. As mentioned earlier, Pashigian et al. uncovered a strong preference for expensive low deductible automobile insurance. Similarly, Robert Eissner and Sotrug (1961) showed that although flight-insurance is considerably more expensive than regular life-insurance, there exists a strong demand for the former. Indeed, flight insurance is one of the few coverages which is actually “bought” (i.e., the consumer rather than insurance agent initiates the purchase).

The above examples suggest that people sometimes ignore low probability events, whereas at other times they focus on the loss dimension. They question whether people are as coherent and consistent as EU maximization implies, highlighting the importance of psychological factors in economic behavior (George Katona, 1975). An interesting illustration of the need to include psychological factors was offered by Katona (1965). He examined whether working people who expect retirement income from pension plans placed more or less funds into private savings (e.g., deposits with banks and savings institutions, bonds or stocks) than working people who are not covered by such pension plans. Private pension plans, which were relatively unimportant in 1945, ex-

---

10 An intriguing economic explanation for this was offered by Sam Peltzman (1975). His data show that the effects of legally mandated installation of various automobile safety devices are offset by increased risk-taking by drivers (e.g., not wearing seatbelts), supporting the notion that drivers determine their own, individually optimal, risk exposure.
panded so much that in 1965 almost one-half of all privately employed workers in the U.S.A. were covered (in addition to almost universal social security coverage). Hence, it presented a unique opportunity for a field study on the effect of private pension plans on discretionary savings, particularly as coverage by these pension plans was largely a function of the type of employment rather than the result of deliberate action to obtain such coverage.

At the outset two alternative hypotheses were formulated. The first derived from traditional economic theory, and suggested that increased "forced savings" (via private pension plans) would reduce voluntary savings, other things being equal. The alternative to this "substitution effect" hypothesis was that increased forced savings would lead to increased voluntary savings because of aspiration level adjustments and goal gradient effects.\(^{11}\) Comparisons between "the average savings ratios of the people covered and people not covered under a private pension plan—paired so as to make them similar in many (socio-economic) characteristics—indicated that the former saved more than the latter" (Katona, 1965, p. 6). These real-world results seem to support an aspiration level and goal gradient hypothesis rather than the traditional economic substitution effect.\(^{12}\)

c. Information Processing Research

One explanation of the above evidence is that people are intendedly rational, but lack the mental capacity to abide by EU theory. This view is corroborated by artificial intelligence research on human perception, recognition, information storage and information retrieval. The bounded rationality view (Simon, 1955) of humans is that of an information processing system which is very narrow in its perception, sequential in its central processing, and severely limited in short-term memory capacity (George Miller, 1956; Simon and Allen Newell, 1971). This limited information processing capacity compels people to simplify even simple problems, and forces them to focus more on certain problem aspects than others (i.e., anchoring). Such adaptation implies sensitivity to the problem presentation, as well as the nature of the response requested. For instance, Joshua Ronen (1973) found that simply interchanging the two stages of a lottery affected preferences (e.g., a .7 chance at a .3 chance of getting $100 was more attractive to subjects than a .3 chance of getting a .7 chance at $100).

Several other studies on gambles also indicate that people adapt their information processing strategy to the specifics of the task. For instance, Slovic and Sarah Lichtenstein (1968a) compared the relative importance of gambles' dimensions using two different response modes. They constructed so-called duplex gambles which contain a separate gain and loss pointer. Such a gamble is fully characterized by four risk-dimensions: namely, a probability of winning ($P_w$), an amount to be won ($A_w$), a probability of losing ($P_l$), and an amount of loss ($A_l$). Using sev-

---

11 The level of aspiration effect, as initially studied by Kurt Lewin (see A. Marrow, 1969, pp. 44–46), states that people raise their sights (i.e., aspirations) with success and lower them with failure. The goal gradient effect refers to a well-known phenomenon in which effort is intensified the closer one gets to a goal.

12 Martin Feldstein questioned this interpretation on grounds of ignoring that "workers who are covered by pensions have an incentive to retire earlier than they otherwise would" (1974, p. 907). He presented time-series data suggesting that public social security depressed personal savings by 30 to 50 percent. Corrections to these analyses by Dan Leimer and Selig Lesnoy (1980), however, found no systematic reduction. A later study by Feldstein (1978) on the effect of private pension programs similarly showed no adverse effects on saving, and possibly an increase. Further evidence bolstering Katona's interpretation derives from a study by Diamond (1977) demonstrating that people do not save enough (by objective standards).
eral such gambles, the following regression model was tested (per subject) $y = \omega + \omega_1 P + \omega_2 A + \omega_3 P + \omega_4 A$, where the dependent variable $y$ represented dollar bids for one group, and rankings on some interval scale for another group of subjects. In the bidding task the dollar dimensions ($A$ and $P$) received significantly more weight than in the ranking task. Apparently subjects adapt their information processing strategies to the response mode: when dollar bids are required dollar dimensions become more salient.

An important issue raised by this study concerns the exact nature of human information processing. Someone might object that Slovic and Lichtenstein (1968a) presupposed a linear regression model (i.e., the above equation) whereas subjects in fact may have processed the moments (i.e., expected value, variance, etc.) of the gamble (as an EU maximizer would). Indeed, several (earlier) studies of gambles support such a moment view, suggesting that expected value and variance are very important (Edwards, 1954b; Coombs and Dean Pruitt, 1960; Lichtenstein, 1965; Norman Anderson and James Shanteau, 1970). The relevant question, therefore, is what results a direct comparison would show.

Two such direct tests were conducted. In another study, Slovic and Lichtenstein (1968b) compared bids between gambles whose stated risk-dimensions (i.e., $P$, $A$, $P$ and $A$) were identical but whose variances were different. The bids for such gambles did not differ significantly (within a pair). Conversely, John Payne and Myron Braunstein (1971) compared equivalent duplex gambles whose stated risk-dimensions differed, but whose moments were identical. This time subjects were not indifferent between the gambles in a pair. Hence, the available data on duplex gambles support the risk-dimension model rather than the moment or EU view. For a review see Payne, 1973. Coombs and Paul Lehner (1981) concluded that the moment model also fails as a descriptor of risk perceptions (as opposed to preference).

Additional evidence for these conclusions derives from the so-called preference reversal phenomenon. Harold Lindman (1971) showed that subjects might prefer gamble A to B when given a direct choice but, at the same time, place a lower reservation price on gamble A than on gamble B when assessed separately. In direct choice subjects focus mostly on the probability of winning, whereas in naming reservation prices the amount of winning is more salient. Such response-induced preference reversals were also demonstrated by Lichtenstein and Slovic (1971) in comparing bids with choices, and have been replicated with real gamblers in Las Vegas (Lichtenstein and Slovic, 1973). Since such reversals may be counter to EU theory or any other holistic choice model, David Grether and Plott (1979) carefully replicated the effect while controlling for such economic explanations as income effects, real vs. hypothetical payoffs, hidden incentives, strategic misrepresentation, etc. None of these hypotheses, however, offered an adequate explanation of the preference reversal phenomenon.

The above results suggest that various choice phenomena cannot be understood or predicted without a detailed understanding of the way information is processed. A natural representation for information processing models are flow-charts, in which components of the alternatives

---

13 For example, if the utility function is linear, only expected value matters. If it is quadratic, both mean and variance are important (H. Levy and Harry Markowitz, 1979). More generally, however, any well-behaved utility function (i.e., one whose derivatives exist and are finite) can be expanded into a Taylor series, whose expected utility is a function of the moments of the gamble (Jack Hirshleifer, 1970; David Baron, 1977).
are compared sequentially, either against each other (Payne and Braunstein, 1971) or against an external reference point. For example, Kunreuther’s 1976 sequential model of flood insurance purchase starts with concern about the hazard. If the probability of loss is below some threshold level, insurance will not be considered seriously, irrespective of the loss and premium (Slovic, et al., 1977).

d. Context Effects

Additional complexities in choice behavior stem from the role of context. The latter refers to the script, verbal labels, social dimensions, information displays, and response modes of a decision problem. Since EU theory focuses on the underlying structure of choices, as modeled by “rational” outside observers, it is largely insensitive to such contextual differences.

Recent studies by Schoemaker and Kunreuther (1979), and Hershey and Schoemaker (1980) uncovered an interesting context effect concerning the wording used in describing decision alternatives. To illustrate, consider the following two formulations for presenting a choice between a sure loss and a probabilistic one:

Gamble formulation:
1a. A sure loss of $10.
1b. A 1 percent chance of losing $1,000.

Insurance formulation:
2a. Pay an insurance premium of $10.
2b. Remain exposed to a hazard of losing $1,000 with a 1 percent chance.

According to EU theory the gamble and insurance formulations involve identical choices between $U(w_b - 10)$ on the one hand, and $[.01U(w_b - 1,000) + .99U(w_a)]$ on the other. Psychologically, however, these two choice situations are quite different. Of 42 subjects, 56 percent preferred the sure loss when presented in the gambling formulation versus 81 percent in the insurance formulation. Hershey and Schoemaker (1980) demonstrated this discrepancy also with other probability and loss levels. The effect was strongest for probability and loss levels representative of insurable hazards (e.g., low probabilities and moderate to large losses). Furthermore, wherever the difference was statistically significant ($p < .05$), the insurance formulation evoked greater risk aversion than did the gamble formulation. Possible reasons for this particular context effect are that different psychological sets are evoked (Robert Abelson, 1976). For example, societal norms about prudent behavior may play a role in the insurance context. Alternatively, different reference points may be used, giving the impression that something is gained in the insurance formulation. Finally, other dimensions may be at play besides money, reflecting the various connotations of insurance such as an assumed administrative cost of filing a claim, less than perfect certainty that repayment will indeed occur in case a legitimate loss is incurred, or regret. However, the first two of these concerns would argue against buying insurance. Regret considerations on the other hand might favor insurance, as the insurance formulation implies that one will know whether or not the loss occurred. Note that this is not implied in the gamble formulation.

Another good example in which isomorphic problems are judged differently was provided by Tversky and Kahneman (1981). Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the consequences are as follows:

A: If program A is adopted exactly 200 will be saved.

B: If program B is adopted, there is a $\frac{1}{6}$ probability that 600 people will be saved, and a $\frac{5}{6}$ probability that no people will be saved.
When this choice was given to 158 subjects, the majority (76 percent) preferred program A. A similar group of 169 subjects were also given the same choice, but in slightly altered form as follows:

A': If program A is adopted exactly 400 people will die.
B': If program B is adopted there is a $\frac{1}{4}$ probability that nobody will die, and $\frac{3}{4}$ probability that 600 people will die.

Although formally equivalent to the earlier formulation, this time only 13 percent preferred program A. This example illustrates how changes in wording can affect the reference point used to evaluate outcomes.

It is a complex issue as to whether context effects are counter to EU theory. Ex post many might fit the model by including additional dimensions, constraints or different reference points. For instance, Bell (1980, 1982) successfully explained such EU anomalies as the Allais paradox, violations of dominance, risk-seeking for losses, preference reversals etc., by introducing regret as the second dimension in a two-attribute expected utility model. Similar ingenious regret explanations were offered by Graham Loomes and Robert Sugden (1981), who suggest that the utility of an object may well be different when it is chosen than when it is received as a gift. Although such EU extensions are intriguing, the EU model does not offer a rich descriptive theory of problem representation and will therefore not easily predict new context effects. The latter requires a better psychological understanding of decision making in general, an issue we turn to next.

V. Psychological Aspects

The failure of EU theory as both a descriptive and predictive model stems from an inadequate recognition of various psychological principles of judgment and choice. Underlying most of these is a general human tendency to seek cognitive simplification. In this section, five psychological aspects are discussed that are particularly germane to risky choice.

First, most decisions are made in decomposed fashion using relative comparisons. Evaluations of multidimensional alternatives are seldom holistic in the sense of each alternative being assigned a separate level of utility. It is cognitively easier to compare alternatives on a piece-meal basis, i.e., one dimension at a time. Two types of such approaches might be distinguished. In the first, alternatives are compared against a preset standard and eliminated if they fail to measure up. If it is required that all attributes meet certain minimum standards, the model is conjunctive. If satisfaction of at least one dimensional standard is sufficient, the model is disjunctive. For example, Melvin W. Reder (1947) found that investment projects are often eliminated from consideration because the probability of ruin exceeds some critical level (i.e., a disjunctive rejection rule). Additional examples are discussed and cited by Roger Shepard (1964), Einhorn (1970, 1971) and Libby and Fishburn (1977).

In the second type, no preset standards are used; instead the alternatives are compared directly (in a decomposed manner). One such approach is elimination-by-aspects, a type of lexicographical model.

In contemplating a dinner at a restaurant, for example, the first aspect selected may be sea food: this eliminates all restaurants that do not serve acceptable sea food. Given the remaining

---

14 This model is non-compensatory in that failure to meet a minimum level on one dimension cannot be compensated for by "surpluses" on others. A good example is the satisficing model associated with Simon’s (1955) theory of bounded rationality. In this model, search among alternatives is terminated once a solution is found that satisfies all preset constraints. Examples of such satisficing behavior in organizational contexts were examined by James March and Simon (1958), Richard Cyert and March (1963), Charles Lindblom (1964) and more recently March and Johan Olsen (1976).
alternatives, another aspect—say a price level—is selected, and all restaurants that exceed the selected price level are eliminated. The process continues until only one restaurant that includes all the selected aspects remains [Tversky, 1972, p. 349].

Another lexicographic model was proposed by Payne and Braunstein (1971) in their study of pairs of duplex gambles. Subjects first compared probabilities. If the difference was sufficiently large a choice would be made on this basis alone. If not, dollar dimensions would be considered. A detailed review of lexicographic models can be found in Fishburn (1974).

An alternative approach is Tversky’s (1969) additive difference model mentioned earlier (p. 542). Empirical evidence for this compensatory model comes from several marketing studies involving binary product choices. J. Edward Russo and Barbara Dosher (1981) used eye-fixation sequences (i.e., whether subjects examined information across rows or columns) as well as verbal protocols to demonstrate that subjects first estimate utility differences within dimensions and then combine these estimates across dimensions. Small differences are often ignored. Similar results were reported by Willem Van Raaij (1977), who examined consumer information-processing strategies under various task structures. The prevalence of such decomposed evaluation strategies raises serious doubts about EU theory’s holistic approach.

A second important principle of choice is that decision strategies vary with task complexity. As Payne (1976) showed, when alternatives and dimensions are numerous, subjects tend to use conjunctive and lexicographic models as initial screening rules. With fewer attributes, holistic evaluations are more probable. However, with just two alternatives and many different attributes the additive difference model is predominant. The idea of variable strategies and mixed scanning was also proposed by Amitai Etzioni (1967), and Irving Janis and Leon Mann (1977). Interesting evidence comes from studies on decision time and task complexity by Charles Kiesler (1966), who examined how much time children took in choosing one piece of candy from four alternatives. Surprisingly, the decision times were shorter when all four pieces were about equally attractive than when two were attractive and two unattractive! This finding suggests a greater motivation to be optimal with simple than with complex choices. Such aspiration shifts are likely to affect decision strategies markedly (Hogarth, 1975b) in ways counter to EU theory.

A third important principle of choice is that of isolation. According to EU theory, decision making requires a portfolio perspective (Markowitz, 1952). For instance, a decision concerning a particular insurance policy would, in theory, depend on all other risks one faces. However, people may not approach decisions in this comprehensive way, and simply treat problems in isolation. Tversky and Kahneman (1981) demonstrated that violations of first-order stochastic dominance can be induced easily because of isolation. Another example they discuss concerns the treatment of a $20 loss. Say that in one scenario you purchase a $20 theater ticket, which you then lose while waiting in the lobby. Would you buy a new ticket? In the other scenario you are about to purchase this same theater ticket, but upon opening your wallet you discover that $20 is missing. Would you still buy the ticket? Because of the way people partition decision contexts, the $20 loss seems less relevant to the second than to the first choice.

---

15 A random variable $\hat{x}$ is said to stochastically dominate another different random variable $\hat{y}$ to the first degree (or order) if the cumulative probability function of $\hat{x}$ nowhere exceeds that of $\hat{y}$. If so, the expected utility of $\hat{x}$ will be higher than that of $\hat{y}$ for any monotonic utility function (Joseph Hadar and William Russell, 1969).
Of course, in both scenarios the choice is either not to go to the theater and be $20 poorer or to go and be $40 poorer. Similar isolation effects are demonstrated in prospect theory (Kahneman and Tversky, 1979) concerning windfall profits, and in Schoemaker (1980) with insurance.

The fourth principle to be mentioned concerns the role of reference points and aspiration levels. Although EU theory suggests that alternatives are evaluated with respect to their effects on final wealth levels, it is cognitively easier to assess options in terms of gains and losses relative to some reference point (Simon, 1955). This target point will often be the status quo, but might also be a target return level, or an aspired future wealth position. Importantly, risk-taking attitudes are likely to be quite different above versus below this reference point. For instance, in prospect theory \( v(x) \) is assumed to be convex for losses and concave for gains (Kahneman and Tversky, 1979). In risk-return models, risk is often related to the probability and consequences of failing to meet a target return level (Fishburn, 1977). Similarly, return may be related to above target probabilities and consequences (Duncan Holthausen, 1981). Although such two-parameter models are descriptively appealing (G. Shackle, 1952), they are theoretically incompatible with EU maximization.

An experimental test of aspiration level effects was offered by Payne et al. (1980, 1981) who examined binary choices of students and managers between three-outcome gambles whose expected values were equal. In each pair, the outcomes of the second gamble all fell within the range of the first. By adding a constant amount to all outcomes, the pairs were translated through the origin to different degrees. Such translations commonly induced changes in preferences, especially if they were such that in one gamble all outcomes were positive (or negative) while in the other gamble there were mixed outcomes. Since at least two general reversals were observed under this design, it would be difficult to explain these results by postulating inflection points in the NM utility function. Instead, these data (especially Payne et al., 1981) point toward strong aspiration level effects. Also see Arthur Williams, 1966.

Another descriptive model that explicitly incorporates the aspiration level concept was proposed by Coombs and Lily Huang (1970). In this so-called portfolio theory, an optimal risk level is assumed for every level of expected value. Gambles of equal expected value are judged in terms of their deviations from this optimal risk level. Hence if two gambles A and B, having equal expected values, are combined into a probability mixture called gamble C, then the new gamble could be closer in risk to the optimal level than either A or B. This model could thus explain the types of the inverted ordering discussed earlier (pp. 542).

An interesting everyday example of inappropriate reference points was offered by Tversky and Kahneman (1981). Suppose you are about to purchase an electronic calculator for $25 and then learn from your friend that exactly the same calculator is for sale at $20 five blocks away. Would you leave the store and go to the cheaper one? Now imagine the same scenario except you are buying a stereo set for $500, and then your friend tells you it is for sale at $495 five blocks away. Would you go this time? Although the crucial question in both cases is whether or not you would walk five blocks for $5, the first saving of 20 percent seems much more attractive than the second which is only 1 percent. Hence, instead of looking at final asset positions, people's reference dimension is percent savings. Pratt, David Wise and Richard Zeckhauser (1979) found field-evidence for this effect in price differences of highly similar
consumer products across stores. The price variance increased markedly with price level, but was fairly constant when expressed as a percentage of price.

Often reference point effects underlie sunk-cost fallacies and failures to treat opportunity costs as being real. As an illustration, consider the following actual example offered by Richard Thaler:

Mr. R bought a case of good wine in the late 50's for about $5 a bottle. A few years later his wine merchant offered to buy the wine back for $100 a bottle. He refused, although he never paid more than $35 for a bottle of wine [1980, p. 43].

Although such behavior could be rationalized through income effects and transaction costs, it probably involves a psychological difference between opportunity costs and actual out-of-pocket costs.

Finally, there is a fifth major factor that underlies EU violations, namely the psychology of probability judgments (Hogarth, 1975a). We shall briefly touch on three aspects: (1) the impact of objective (or stated) probabilities on choice, (2) quantitative expressions of degrees of beliefs through subjective probabilities, and (3) probabilistic reasoning, particularly inference.

A general finding of subjective expected utility research is that subjective probabilities relate non-linearly to objective ones (e.g., Edwards, 1953; 1954a). Typically, the subjective probability curves overweigh low probabilities and underweigh high ones (Wayne Lee, 1971, p. 61). For example, Menahem Yaari (1965) explained the acceptance of actuarially unfair gambles as resulting from optimism regarding low-probability events, as opposed to convexities in the utility function. Although the evidence is inconclusive as to the general nature of probability transformations, particularly with respect to its stability across tasks and its dependence on outcomes (Richard Rosett, 1971), R. W. Marks, 1951, Francis Irwin, 1953, and Slovic, 1966 found that subjective probabilities tend to be higher as outcomes become more desirable (i.e., a type of wishful thinking). Furthermore, subjective "probabilities" often violate mathematical properties of probability (Kahneman and Tversky, 1979). In such cases we refer to them as decision weights $w(p)$, as noted earlier.

In expressing degrees of confidence probabilistically, other biases are encountered. A robust result is that people are generally overconfident. For instance, if one were to check out all instances where a person claimed 90 percent confidence about the occurrence of events, typically that person would be correct only about 80 percent of the time (Lichtenstein, et al., 1981). This same bias manifests itself in subjective confidence intervals, which are typically too tight (M. Alpert and Raiffa, 1981). Many biases in subjective probability judgments stem from the type of heuristics (i.e., simplifying rules) people use to estimate relative likelihoods. For instance, a doctor may judge the likelihood of a patient having disease A versus B solely on the basis of the similarity of the symptoms to textbook stereotypes of these diseases. This so-called representativeness heuristic, however, ignores possible differences in the a priori probabilities of these diseases (Kahneman and Tversky, 1972). Another common heuristic is estimation on the basis of availability (Tversky and Kahneman, 1973). In judging the chances of dying from a car accident versus lung cancer, people may base their estimates solely on the frequencies with which they hear of them. Due to unrepresentative news coverage in favor of car accidents, these estimates will be systematically biased (Lichtenstein et al., 1978). Finally, people suffer from a serious hindsight bias resulting in suboptimal learning. Events that did (not) happen appear in retrospect more (less) likely than they did before the outcome was known. Baruch Fischhoff (1975) explains this phenome-
non as due to reconstructive memory. Once new information is received it is combined with older information, making the latter irretrievable by itself.

A third area where probabilities are improperly estimated is Bayesian inference in well-structured tasks. Two somewhat opposite phenomena are encountered depending on the context. One is conservatism (Edwards, 1968) according to which new information is underweighted in the revision of opinions (e.g., in poker chip experiments). It reflects an anchoring onto the old information with insufficient adjustment in the direction of the new information. The other phenomenon involves ignorance of stated prior probabilities, similar to the earlier example of disease A vs. B. One psychological explanation is that people’s heuristic for assessing the relevance of information is based on its causal relationship (Tversky and Kahneman, 1980). This would explain why most people consider the probability of a daughter having blue eyes, given that her mother has blue eyes, to be higher than the probability of a mother having blue eyes, given that her daughter has blue eyes. However, assuming that successive generations have the same incidence of blue eyes, these probabilities are the same.

Perceptions regarding causal connections may similarly explain people’s general insensitivity to prior probabilities, the causality of which is often unclear. For instance, in judging the probability that a taxi involved in some city accident was blue rather than green, subjects focused entirely on the reliability of the eye-witness testimony while ignoring the explicitly stated information that only 10 percent of the taxis in the city were blue. However, when Bar-Hillel (1980) rephrased the problem by saying that although there were as many blue as green cabs in the city, only 10% of the taxis involved in past accidents were blue, this prior information received considerable weight. Even though this problem is statistically identical to the earlier version, emphasizing the causal connection of the prior probabilities to the event of interest markedly improved subjects’ posterior probabilities (as judged from Bayes’ theorem). The above examples underscore that subjective probabilities commonly violate basic statistical principles, thereby invalidating axiom five in NM utility theory as well as Savage’s SEU axioms.

VI. Conclusions

The research reviewed in this article suggests that at the individual level EU maximization is more the exception than the rule, at least for the type of decision tasks examined. To assess the role of EU theory more generally, e.g., for future decision models, we next examine the evidence from each of the four perspectives identified at the beginning of the paper.

As a descriptive model seeking insight into how decisions are made, EU theory fails on at least three counts. First, people do not structure problems as holistically and comprehensively as EU theory suggests. Second they do not process information, especially probabilities, according to the EU rule. Finally, EU theory, as an “as if” model, poorly predicts choice behavior in laboratory situations. Hence, it is doubtful that the EU theory should or could serve as a general descriptive model. However, there may be exceptions. For well-structured repetitive tasks, with important stakes, and well-trained decision makers, EU maximization may well describe the actual decision process, e.g., oil-drilling decisions. Indeed, in large organizations where computers are used and highly trained managers operate, the EU model might be explicitly followed. However, even under such favorable circumstances, problem definitions and solutions can be plagued by sunk-cost fallacies, isolation effects, and asymmetrical treat-
ments of opportunity and out-of-pocket costs. As Thaler (1980) argues, these biases afflict organizational as well as individual decision making.

From a positivistic perspective, the interpretation of evidence counter to EU theory is more complicated. A well-known example of Friedman and Savage (1948) notes that complex equations of rigid body mechanics and plane geometry may offer excellent predictions of the manner in which expert billiard players take their shots, even if these players are totally ignorant of such equations. The reason is that the geometric model captures well what the players try to do. Since they are experts, years of training and feedback have led to heuristics that closely approximate optimal behavior. Four conditions, however, make this analogy of limited value to economic behavior. First, most people are not experts in economic matters (Thaler, 1980). Second, learning from feedback is not a simple or automatic activity in daily decision making. Uncertainty, environmental instability, improper assessment frameworks, and lack of insight into one's decision rules are all serious obstacles to learning from experience (Einhorn, 1980). Third, the optimality of economic behavior in real-world settings is often difficult to assess without specific knowledge of the person's utility function, the particular problem perception, and the rationality criteria being pursued (March, 1978). Finally, even if the EU model did predict well (while its assumptions are wrong) the notion that only prediction matters is epistemologically quite unappealing (Paul Samuelson, 1963). For instance, it would be interesting to know what makes the model so robust under misspecification (Robyn Dawes, 1979).

A second reaction, from a positivistic economic perspective, is to note that individual biases and differences may not matter in aggregate behavior. For example, Stigler and Becker (1977) argued that people may usefully be treated as similar in basic tastes and preferences, both across time and persons. Moreover, market mechanisms may be presumed to correct misperceptions and individual decision biases. However, not necessarily. Paul Kleindorfer and Kunreuther (1982) examined analytically the effect of probability misinformation (by firms and consumers) on insurance markets. Using Rothschild's and Joseph Stiglitz' (1976) solution to the adverse selection problem (i.e., allowing insurers to offer varying dollar premiums as a function of total coverage), they established various conditions under which the market fails. Pratt, Wise and Zeckhauser's 1979 study on price differences among highly similar consumer products also demonstrates a failure of the market to correct individual biases. Finally, Thomas C. Schelling (1978) discusses many other examples of "irrational" macro behavior, such as drivers incurring a thirty minute delay to have a one-minute peek at some highway accident. The point to be made is that the connection between micro and macro behavior is too complex to argue that at higher levels of social aggregation individual biases generally wash out or self-correct.

As a third discussion point, let us examine the view that hypothetical and "artificial" laboratory experiments have limited implications for economic theory. Sociological research (H. Schuman and M. P. Johnson, 1976) indeed underscores the consideration that intentions (i.e., hypothetical choices) cannot be mechanically substituted for, or assumed to be generally highly correlated with, actual behavior. In the context of gambles, however, several empirical studies by Nathan Kogan and Michael Wallach (1967) and Slovic (1969) on the effect of hypothetical vs. real payoffs suggest that cognitive processes and decisions do not differ considerably between these two conditions. Indeed, I know of no evidence that suboptimal labo-
ratory behavior improves when committing subjects financially to their decisions. For example, Lichtenstein and Slovic (1973) found similar biases and inconsistencies in their Las Vegas replications with real gamblers (who played with their own money for large stakes), as they had earlier found with college students who had provided hypothetical judgments (Lichtenstein and Slovic, 1971). Similarly, Grether and Plott's (1979) recent replication of the so-called preference reversal phenomenon, using gambles with real money, revealed similar, and at times greater inconsistencies than under the hypothetical condition. From these studies, it appears unlikely that the subjects in the experiments discussed would come closer to optimal EU behavior when making decisions for real. The failure to optimize appears to be cognitive (i.e., related to the way problems are structured and what decision strategies are used) rather than motivational (i.e., the amount of mental effort expended). This brings us to the charge that laboratory experiments tend to be artificial, which is confused here with being unrepresentative (Karl Weick, 1967). Behavior in the laboratory is as real as other forms of behavior. If economic theory is proposed as a general model of scarce resource allocation, it should apply to experimental settings as well (Vernon Smith, 1976). Indeed, the burden of proof should be on those wishing to exclude laboratory behavior from economic theory rather than on those who want to include it.

From a postdictive perspective, the EU violations discussed would all be considered illusory. It would be argued that the examples cited appear to violate the EU model because of improperly specified costs and benefits. The ensuing rationalization would focus on hidden costs, incentives, dimensions and constraints. For instance, intangible cognitive costs might be introduced, even with justification (e.g., Steven Shugan, 1980). However, to mount an elaborate rescuing mission along this line may be self-defeating. The postdictive view is particularly prone to fall prey to a methodological sunk-cost fallacy. Having invested heavily in complex deductive structures, with wide domains of applicability and mathematically elegant decision models that allow for easy aggregation across people, it is a natural tendency to patch up the theory cosmetically. A better alternative is to examine closely the type of anomalies reported, and the cognitive reasons underlying them. Although cognitively realistic choice models may be task dependent and presented in process or flow-chart form, they can often be approximated holistically (Einhorn et al., 1979). For instance, a chess program that provides a cognitive representation of some human chess master might be approximated by a set of functions that are to be maximized. Once such optimality-based approximations are obtained they can be connected with traditional economic theory. The point is that modifications to economic theory (while retaining some sort of optimality) should be based on cognitive insights (see Roy Lachman, Janet Lachman, and Earl Butterfield, 1979 for a review) rather than ad hoc rationalization or mathematical expedience.

Finally, there is the prescriptive or normative perspective. At first glance, the numerous EU violations cited strengthen the case for formal decision analysis as a way of circumventing and supplementing the suboptimal nature of intuitive, unaided decision making. However, there are two related implications we should consider. First, some of the biases described (particularly regarding the axioms) may be so basic that they render the normative theory inoperational. Second, persistent violations of the EU model may raise questions as to its normative validity.

Regarding the operationality issue, an important difficulty concerns the con-
struction of NM utility functions. As Cornelius Van Dam (1973) and Karmarkar (1978) have shown, 50–50 reference lotteries will often lead to a different NM utility function than, for example, would result if 30–70 lotteries had been used. Furthermore, as shown in Hershey and Schoemaker (1980) and Tversky and Kahneman (1981), subtle changes in the context or framing of a problem may lead to different preferences (see also Don Wehrung et al., 1980). Thus the question arises which context measures the “true” risk-taking attitude (Jim Barnes and James Reinmuth, 1976, and Hans Binswanger, 1980) or more fundamentally, whether there really exist basic tastes and preferences that are compatible with the EU axioms. A recent study by Hershey, Kunreuther and Schoemaker (1982), focusing on response mode biases, probability distortions, translation effects, risk transfer asymmetries and context effects, suggests that the answer to the latter question is no.16

One solution many normative theorists (e.g., Keeney and Raiffa, 1976) would recommend is to explain to the decision maker his or her inconsistencies and see if a revision of preference(s) is desired. For instance, MacCrimmon (1968) found that executive subjects would often change their choices when made aware of violations of normative postulates (e.g., those of Savage, 1954). However, social pressure and conformity tendencies may have confounded these results. In a follow-up study, Slovic and Tversky (1974) controlled for these influences. As mentioned earlier, they tested Savage’s independence principle by presenting subjects with the Allais (1953) and Ellsberg (1961) paradoxes. Once subjects had made their choices, they were then given a prepared, authoritative argument against their particular choice (i.e., either Allais’ position or that of Savage). After reflecting on these arguments, the subjects were subsequently asked to reconsider their choice. Except for a few individuals, most subjects did not change their preferences, many of which were in violation of Savage’s independence principle. Indeed, even among experts these paradoxes evoked considerable debate concerning the normative acceptability of the EU postulates (e.g., Samuelson, 1952; Allais, 1953; Savage, 1954; Raiffa, 1961; and Ellsberg, 1961 and 1963), which brings us to the second normative implication.

As MacCrimmon and Larsson (1979) noted: “since many careful, intelligent decision makers do seem to violate some axioms of expected utility theory, even upon reflection of their choices, it does seem worthwhile exploring this third option of considering modifications of the standard theory” (p. 83). In this vein Chew and MacCrimmon (1979a and 1979b) recently proposed an alternative generalized EU model, in which the substitution axiom is considerably weakened. Mark Machina (1982) showed that the major economic concepts and tools of EU analysis do not depend on the independence axiom. As an alternative route, Loones and Sugden (1981) proposed to drop the transitivity axiom, both descriptively and normatively. Finally Fishburn (1981) recently developed an alternative preference representation that utilizes neither the transitivity nor the independence axiom.

Apart from modifications to the normative model two other options were presented by MacCrimmon and Larsson (1979), namely: (1) to maintain that one’s choices are valid and that the assumptions or axioms do not apply in the given case, or (2) to change one’s choices to conform

---

16 Recent experiments suggest that the mathematical EU form of separable \( f(p) \) and \( U(x) \) transformations is questionable (John Lynch, 1979 and Lehner, 1980). Rather than reflecting risk-attitude indirectly in \( U(x) \), it is proposed that the utility function explicitly incorporates probability, i.e., \( EU = \sum f(p_k) U(x_i, p_k) \).
with the axioms. Let us examine the former, which can be interpreted in two ways: (1) someone disagrees with the probabilities used in showing violations of EU, or (2) the notion of objective risk is rejected, and hence the problem representation is disputed. Regarding the first one, objective probabilities exist only in hypothetical situations (and even this is disputable). In reality they are estimated and subjective. Since legitimate differences may exist among people in their subjective probabilities for the same event, violations of EU may be difficult to prove other than in hypothetical choice situations.

More basically, however, someone could refute a particular problem representation. For example, in the previously discussed context effect (Hershey and Schoemaker, 1980), the insurance formulation elicited more risk averse behavior than did the gamble formulation. However, this apparent EU violation might be explained by assuming that the outcome space is different in the insurance formulation. Since problem representation is inherently a subjective matter, it is subject to only limited normative evaluation (L. Jonathan Cohen, 1981). Indeed, there exists no general normative theory as to how problems should be defined, or how language and context should be encoded.\textsuperscript{17} Moreover, someone might fundamentally question the meaningfulness of the notions of probability and risk. As noted earlier, there exist difficult, unresolved philosophical problems in the area of probability (Arthur Burks, 1977). To quote West Churchman (1961, p. 139): “almost everyone knows what it means to say that an event is only probable—except those who have devoted their lives to thinking about the matter.” Historically, probability is a relatively recent formal construct (Ian Hacking, 1975), which lacks primary sensory evidence as to its existence. As such it might be viewed as an invention rather than a discovery. Note in this regard the finding by Lawrence Phillips and C. N. Wright (1977) that the Chinese are less likely than the English to view the world probabilistically. Hence, an extreme but tenable attitude is to view the EU model as an interesting theoretical construction which is useless for real-world decision-making.

In conclusion, we have attempted to review recent evidence concerning the EU model from four different perspectives. Although the evidence and associated interpretations have been critical as to the model’s usefulness, it must be emphasized that much of the research would not have resulted without the existence of EU theory in the first place. As such, the model has yielded deeper insights and more refined questions, both descriptively and normatively, concerning decisions under risk. It has revealed that people perceive and solve problems differently, and has offered a framework and language in which to discuss these differences. Our intellectual indebtedness to the EU model is thus great, although its present paradigmatic status (in certain fields) should be questioned. Nevertheless, until richer models of rationality emerge, EU maximization may well remain a worthwhile benchmark against which to compare, and toward which to direct, behavior. On the other hand, it is likely that today’s paradoxes and persistent EU violations hold the seed of future normative as well as descriptive theories of choice. After all it was a paradox (Bernoulli, 1738) that gave birth to the current normative model.

\textsuperscript{17} Churchman (1971) discusses various philosophical approaches to defining rationality and structuring inquiry systems concerning the nature of the world. A strong prior commitment to axiom systems, as in the EU model, is a Leibnizian approach. If the formal model is correct, this strategy would be highly effective; if wrong, however, its costs could be considerable due to delayed detection of miscalculations. A more prudent strategy is the dialectical approach of Hegel or Kant (see Richard Mason and Ian Mitroff, 1973).
REFERENCES


BLAUG, MARK. The methodology of economics. N.Y.: Cambridge Univ. Press, 1980.


AND MACCRIMMON, KENNETH R. “Alpha Utility Theory, Lottery Composition, and the Allais Paradox.” Working Paper, Faculty of Com-
merce and Business Administration, Univ. of Brit-
ish Columbia, Vancouver, Canada, Sept. 1979b.

CHIPMAN, JOHN S. "The Foundations of Utility," Eco-

CHURCH, C. WEST. Prediction and optimal deci-
sion: philosophical issues of a science of values.

—. The design of inquiring systems; basic con-
cepts of systems and organization. New York: Ba-

CODY, M. "Optimization in Ecology," Science, 1974,
183, pp. 1156–64.

COHEN, L. JONATHAN. "Can Human Irrationality Be
Experimentally Demonstrated?" The Behav. &

COOMBS, CLYDE H. "Portfolio Theory and the Mea-
surement of Risk," in Human judgment and deci-
sion processes. Edited by MARTIN F. KAPLAN AND
STEVEN SCHWARTZ. New York: Academic Press,
1975, pp. 63–86.

— AND HUANG, LILY C. "Tests of a Portfolio

— AND LEHNER, PAUL E. "An Evaluation of
Two Alternative Models for a Theory of Risk," 
Working Paper, Dept. of Psychology, Univ. of
Michigan, Ann Arbor, 1981.

— AND PRUITT, DEAN G. "Components of Risk
in Decision Making: Probability and Variance
77.

Cramer, Gabriel. Letter to Nicolas Bernoulli, a
cousin of Daniel (see Bernoulli 1738a, above).

Cyert, RICHARD M. AND MARCH, JAMES G. A be-
havioral theory of the firm. Englewood Cliffs, NJ:

Davidson, Donald and Suppes, Patrick. "A Fin-
istic Axiomatization of Subjective Probability and

— AND SIEGEL, SIDNEY. Decision making: an ex-
perimental approach. Stanford, Calif.: Stanford
Univ. Press, 1957.

Dawes, Robyn M. "The Robust Beauty of Improper
Linear models in Decision Making," Amer. Psych.

Debreu, Gerard. Theory of value: An axiomatic
analysis of economic equilibrium. New York: Wi-
ley, 1959.

De Finetti, Bruno. "La Prévision: Ses Lois Logi-
ques, Ses Sources Subjectives," Annales de l’In-
stitut Poincaré, 1937, 7, pp. 1–68.

—. Theory of Probability: A Critical Introduc-
tory Treatment. Vol. 1. Translated by ANTONIO
MACHI AND ADRIAN SMITH. N.Y. Wiley [1970],
1974.

Diamond, Peter A. "A Framework for Social Secu-
rity Analysis," J. Public Econ., Dec. 1977, 8(3),
275–98.

— AND ROTHSCCHILD, MICHAEL. Uncertainty in
economics: Readings and exercises. New York: Ac-

Drèze, Jacques H. "Axiomatic Theories of Choice,
Cardinal Utility and Subjective Probability: A Re-
view," Allocation under uncertainty: Equilibrium
and optimality. Edited by Jacques H. Drèze.

Dyer, James S. and Sarin, Rakesh K. "Measurable

— AND SARIN, Rakesh K. "Group Preference
Aggregation Rules Based on Strength of Prefer-
ence," Manage. Sci. 25(9), 1979b, pp. 822–32.

— AND SARIN, Rakesh K. "Relative Risk Aver-

Edwards, Ward. "Probability-PREFERENCES in Gam-

—. "Probability Preferences Among Bets with

—. "Variance Preference in Gambling," Amer.

—. "The Prediction of Decisions Among Bets,"


—. "Conservatism in Human Information Pro-
cessing," Formal representation of human judg-
ment. Edited by Benjamin Kleinnuntz. New

Ehrlich, Isaac and Gary S. Becker. "Market In-
surance, Self-Insurance, and Self-Protection," J.

Einhorn, Hilllel J. "The Use of Nonlinear, Non-
compensatory Models in Decision Making," Psych.

—. "Use of Nonlinear, Noncompensatory Models
as a Function of Task and Amount of Information,"

—. "Learning From Experience and Suboptimal
Rules in Decision Making," in Cognitive processes
in choice and decision behavior. Edited by
Thomas S. Wallsten. Hillsdale, NJ: L. Erlbaum, 1980,

— AND HOgARTH, Robin M. "Behavioral Deci-
sion Theory: Processes of Judgment and Choice,"

—; Kleinnuntz, Don H. and Kleinnuntz, Ben-
jamin. "Linear Regression and Process-Trac-
pp. 465–85.

Eisenb, Robert and Strotz, Robert H. "Flight

Ellsberg, Daniel. "Classic and Current Notions of
'Measurable Utility'," Econ. J. 64, Sept. 1954,
pp. 528–56.

—. "Risk, Ambiguity and the Savage Axioms,"

—. "Risk, Ambiguity and the Savage Axioms:
41.

Etzioni, Amitai. "Mixed Scanning: A Third Ap-
proach to Decision-Making," Public Admin. Rev.,

Fama, Eugene F. "Ordinal and Measurable Utility,"
Studies in the theory of capital markets. Edited
45.
_____.


_____.


_____.


_____.


_____.


_____.


_____.


_____.


_____.


Kiesler, Charles A. “Conflict and Number of


**LINDSAY, R. B.** “Physics—To What Extent is it Deterministic?,” *Amer. Scientist*, 1968, 56, pp. 93–111.


MILLER, GEORGE A. “The Magical Number Seven, Plus or Minus Two; Some Limits on our Capacity for Processing Information,” Psych. Rev. 1956, 63 (2), pp. 81–97.


SLOVIC, PAUL. “Value as a Determiner of Subjective Probability,” IEEE Transactions on Human Fac-


