ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 44, 166–202 (1989)

Parametric Models of the Utility of Survival Duration Tests of Axioms in a Generic Utility Framework

JOHN M. MIYAMOTO

University of Washington

AND

STEPHEN A. ERAKER

Department of Internal Medicine, Kaiser Permanente, Milpitas, California

The risk posture of a utility function is, roughly speaking, a measure of the curvature of the utility function with respect to an objectively quantified dimension like money or survival duration. Risk posture with respect to survival duration is critically important in the normative decision analysis of medical therapy selection because medical choices often involve therapies that differ in their trade-offs between short- and long-term survival, e.g., between surgical and nonsurgical treatments. Power utility functions and exponential utility functions are potentially useful in the normative analysis of medical decisions because they provide simple representations of risk posture with respect to survival duration. The present study formulates and tests axiomatic characterizations of power and exponential utility models within a theoretical framework called the generic utility theory. The formalizations generalize the work of J. W. Pratt on parametric characterizations of risk attitude. An experiment is reported in which the hypotheses that the utility of survival duration is a linear, exponential, logarithmic, or power function are tested. The experiment illustrates how to apply statistical criteria to the empirical test of preference axioms. Experimental results show that the linear, exponential, logarithmic, and power utility models are all violated by a substantial proportion of subjects. Qualitative analyses of the pattern of violations suggest that the linear and exponential utility models are a better approximation of subjects' preference than the logarithmic and power utility models. The formalizations and experimental tests are carried out within the generic utility theory because the assumptions of this theory are consistent with important "strong" utility theories, including expected utility theory, subjective expected utility theory, the dual bilinear model, and prospect theory. Investigations of utility models within the generic utility framework are interpretable from the standpoint of stronger theories, but are not committed to the hypothesis that a particular strong utility theory is valid.

We gratefully acknowledge the advice and criticism of Clyde Coombs, Richard Cornell, David H. Krantz, and J. Frank Yates. We would like to thank Diane Bushberg for conducting the interviews of half of the subjects in this experiment. A preliminary version of the present experiment was reported in a doctoral dissertation by the first author. Requests for reprints should be sent to John M. Miyamoto, Department of Psychology, University of Washington, Seattle, WA 98195.

0749-5978/89 $3.00

The utility of survival duration is critically important in the normative decision analysis of medical therapy selection. In particular, the risk posture of the utility function, which is roughly speaking the degree to which it is concave or convex, affects the decision analysis of therapeutic choices when the therapies under consideration differ in their trade-offs between short- and long-term survival (McNeil & Pauker, 1982; Pauker & McNeil, 1981). A striking illustration of this point is found in the analysis of treatment selection in lung cancer (McNeil, Weichselbaum, & Pauker, 1978). Surgical treatment of lung cancer confers a greater chance of long-term survival, but chances for short-term survival are inferior due to operative mortality. Radiation treatment of lung cancer provides a superior chance of short-term survival because it has no operative mortality, but long-term survival is inferior. The survival curves for the therapies cross at about 2 years. McNeil *et al.*, (1978) showed that the expected utility of radiation therapy tends to be higher if the utility of survival duration is risk averse (concave downward), whereas the expected utility of surgical treatment tends to be higher if the utility of survival duration is risk seeking (concave upward).

The precise demarcation between the utility functions that are optimized by radiation and surgical treatment is determined by the surgical curves of the patient when he or she receives that therapy, and these in turn are determined by a more detailed specification of the health status of the patient. Nevertheless, the McNeil *et al.* (1978) study illustrates an important qualitative principle: if a therapeutic choice involves trade-offs between chances for short- and long-term survival, risk-seeking utility functions tend to be more favorable to the choice with a higher chance of long-term survival, and risk-averse utility functions tend to be more favorable to the choice with a higher chance of short-term survival. Surgical versus nonsurgical choices often have this character, and hence, are sensitive to risk posture with respect to survival duration (McNeil & Pauker, 1982; Pauker & McNeil, 1981).

Henceforth when we refer to risk posture, we mean risk posture with respect to survival duration (unless we specify otherwise). Although risk posture usually refers in utility theory to risk posture with respect to money, the restriction of the term to survival duration is reasonable because that is the subject of these investigations. We should also point out that the theoretical models and experimental methods developed here are equally applicable to risk posture with respect to money or any other quantified commodity or attribute.

The present study focuses on the issue of whether risk posture can be characterized by one of several simple classes of parametric utility models. In particular, we will investigate whether the utility of survival duration can be described by linear, exponential, logarithmic, or power functions of duration. To motivate this choice of research problem, we will briefly describe a health utility model due to Pliskin, Shepard, and Weinstein (1980) that uses power functions to represent risk posture and a multiplicative factor to represent the effect of health quality on the utility of survival. According to the model of Pliskin *et al.*, the utility of surviving y years in a health state q is given by the equation

$$U(y, q) = y^{\omega}H(q) \tag{1}$$

In Eq. (1), $U(Y, q)$ is the utility of the (Y, q) outcome, ω is a positive parameter, and H is a function that assigns utilities to qualities. Model (1) makes the simplifying assumption that a single health state q describes health quality during the y years of survival. A more realistic model would consider the possibility that a succession of different health states, q_1, q_2, q_3, . . . , are experienced during survival. Although the investigation of the more realistic model is an ultimate goal of health utility analysis, studies of simple models like (1) are important steps toward this goal (Miyamoto, 1987b; Miyamoto & Eraker, 1985, 1988; Pliskin *et al.*, 1980; Weinstein, Plisken, & Stason, 1977; Weinstein & Stason, 1982). Model (1) embodies two claims. First, when health quality is held fixed, the utility of y years is a power function of y. In the technical terminology of utility theory, the utility function is risk neutral (linear) when $\omega = 1$, risk averse when $\omega < 1$, and risk seeking when $\omega > 1$ (Miyamoto & Eraker, 1985). The value of ω determines an individual's propensity to trade chances for short-term and long-term survival. Second, model (1) asserts that the utility of surviving y years is multiplied by a factor $H(q)$, which represents the relative worth of surviving in health state q. For example, suppose that q^* designates whatever health state is regarded as most desirable, and let us arbitrarily specify that $H(q^*) = 1$. Then $H(q) = y^{\omega}H(q)/y^{\omega}H(q^*) = U(y, q)U(y, q^*)$. Thus, $H(q)$ can be interpreted as the worth of y years of survival in health state q relative to the worth of y years in the most desirable health state q^* (Miyamoto & Eraker, 1985, 1988; Pliskin *et al.*, 1980).

Miyamoto and Eraker (1985) estimated values of ω and $H(q)$ in a sample of 46 coronary artery disease patients. In this study, the levels of q represented survival with and without chest pain (*angina pectoris*). Estimated values of ω ranged from .23 and 12.95. Setting H(freedom from angina) = 1, the value of H for survival with existing levels of angina ranged from .17 to 1.00. Miyamoto and Eraker did not evaluate what therapeutic choice would optimize the utility of the patients because they did not have estimates of survival and health quality probabilities for the individual patients, but they pointed out that the expected utility of surgical treatment of coronary artery disease increases relative to that of medical treatment as values of ω increase and values of H for survival with angina decrease. Weinstein and Stason (1982) analyzed the cost-effectiveness of coronary artery bypass graft surgery for different categories of patients. Although their study was not based on empirically assessed utility functions for survival duration and health quality, they used model (1) to evaluate the expected utility of surgical treatment and medical management of coronary artery disease when ω and $H(q)$ were assumed to take on various values.[1]

These examples are cited here to show that a utility model like (1) is useful because the values of ω and $H(q)$ provide simple summary measures of risk posture and the relative worth of health qualities, which are critical variables in decision analyses of treatment selection. Later, we will discuss more general forms of power utility models, and also exponential utility models, and show how to formalize and test the empirical validity of such models. The important point here is that the classes of power and exponential utility models both provide simple characterizations of risk posture which is known to be critically important in medical decision analysis.

The issue of how health quality affects the utility of survival duration will not be addressed in the present study. Although one is naturally interested in the question of how health quality affects the utility of survival duration, there are several reasons to focus on duration alone. First, in some medical decision analyses, the therapeutic alternatives under consideration yield similar health qualities so that duration is the primary criterion of choice (McNeil *et al.*, 1978). Second, if a multiplicative model describes the utility of survival duration and health quality, as some have suggested (Miyamoto, 1985, 1987b; Miyamoto & Eraker, 1985, 1988; Pliskin *et al.*, 1980), risk posture in the technical sense of multiattribute utility theory is independent of health quality. In the technical sense, risk posture refers to only those features of the utility function that affect preferences between gambles (Keeney & Raiffa, 1976; Pratt, 1964; Von Winterfeldt & Edwards, 1986); as is well known, a multiplicative model predicts that preferences between gambles for survival duration in a fixed health state are unaffected by the choice of the fixed health state (Keeney & Raiffa, 1976; Miyamoto & Eraker, 1988; Pliskin *et al.*, 1980; Von Winterfeldt & Edwards, 1986). Hence, if health quality and survival duration satisfy a multiplicative utility model, a description of risk posture in any one health state is also a description of risk posture in any other health

[1] A number of other medical decision analyses have used model (1) under the restrictive assumption that $\omega = 1$ (risk neutrality) (cf. Pliskin *et al.*, 1981; Weinstein, Pliskin, & Stason, 1977).

state. Third, even if the utility of health quality and survival duration is not multiplicative, the analysis of the more general problem will be advanced by a better understanding of risk posture with respect to survival in a fixed health state.

Generic Utility Theory

Parametric utility models will be investigated here in a theoretical framework called generic utility theory (Miyamoto, 1988). We cannot give a complete exposition of the theory here, but it will be helpful to explain why the investigations are conducted within this theoretical framework. As is well known, expected utility (EU) theory and subjective expected utility (SEU) theory, which were at one time the dominant theories of preference under risk, have been shown experimentally to be violated by actual preferences of individuals (Grether & Plott, 1979; Kahneman & Tversky, 1979; Slovic & Lichtenstein, 1983; Von Winterfeldt & Edwards, 1986). A number of alternative theories of preference under risk have been developed to account for actual preference behavior, but the evidence does not yet point uniquely to one theory over all competitors. The generic utility theory is a weakly constrained theory in the sense that its assumptions are compatible with a number of stronger theories including EU and SEU theories, Edwards' (1962) additive and nonadditive subjective expected utility (ASEU and NASEU) theories, Kahneman and Tversky's (1979) prospect theory, Karmarkar's (1978) subjectively weighted utility (SWU) model, and Luce and Narens' (1985) dual bilinear model. The advantage of the generic utility framework is that formalizations and experimental tests within this framework are also interpretable from the standpoint of these stronger theories. In view of current conflicts at the foundations of preference theory, it is useful to have a framework that is just strong enough to permit the formalization and testing of utility models without committing one to the full set of assumptions underlying a strong theory like SEU or prospect theory.

We will first state the generic utility theory, and then show that it is possible to formulate and carry out rigorous experimental tests of the linear, exponential, logarithmic, and power utility models in the generic utility framework. The presentation of the generic utility theory will be restricted to aspects of the theory that are needed to formulate tests of the parametric utility models. We will develop the theory in terms of the utility of survival duration, although other commodities or attributes would also serve to illustrate the theory.

Suppose that x and y are possible survival durations (nonnegative durations that are not unreasonably long). Let p be a probability not equal to zero or one, and let $(x\,p\,y)$ denote a gamble with a p chance of receiving x, and a $1 - p$ chance of receiving y. By a p-gamble, we will mean any

gamble of the form $(x\,p\,y)$. We will assume that some particular probability p has been chosen, and that this choice remains fixed throughout the analysis. Let $x \succcurlyeq_{\mathrm{p}} y$ indicate that x is equal or more preferred than y, and let $(x\,p\,y) \succcurlyeq_{\mathrm{p}} (w\,p\,z)$ indicate that the gamble $(x\,p\,y)$ is equal or more preferred than the gamble $(w\,p\,z)$. (The subscript p on the $\succcurlyeq_{\mathrm{p}}$ relation refers to preference, and not the probability p.) The distinctive feature of the generic utility theory is that the utility representation is only posited to hold on certain subsets of p-gambles, defined as follows. A set G of p-gambles will be called a *lower triangular set* if and only if $x \succcurlyeq_{\mathrm{p}} y$ for every $(x\,p\,y)$ in G. A set G' of p-gambles will be called an *upper triangular set* if and only if $y \succcurlyeq_{\mathrm{p}} x$ for every $(x\,p\,y)$ in G'. The rationale for the terminology is that if p-gambles were arranged on the $X \times Y$ plane, with (x, y) being the coordinates of $(x\,p\,y)$, the lower triangular gambles would be on or below the diagonal, and the upper triangular gambles would be on or above the diagonal.

Suppose that G is a lower triangular set of p-gambles. G will be said to have a generic utility representation provided that there exists an interval scale U and positive constants α and β that satisfy

$$(x\,p\,y) \succcurlyeq_{\mathrm{p}} (w\,p\,z) \quad \text{iff } \alpha U(x) + \beta U(y) \geqslant \alpha U(w) + \beta U(z) \tag{2}$$

for every $(x\,p\,y)$ and $(w\,p\,z)$ in G ("iff" abbreviations "if and only if"). Similarly, suppose that G' is an upper triangular set of p-gambles. G' will be said to have a generic utility representation provided that there exists an interval scale U and positive constants α' and β' that satisfy

$$(x\,p\,y) \succcurlyeq_{\mathrm{p}} (w\,p\,z) \quad \text{iff } \alpha' U(x) + \beta' U(y) \geqslant \alpha' U(w) + \beta' U(z) \tag{3}$$

for every $(x\,p\,y)$ and $(w\,p\,z)$ in G'. Although the utility scale U is assumed to be the same in (2) and (3), the coefficients α, α', β, and β', need not satisfy $\alpha = \alpha'$ and $\beta = \beta'$. We will assume that $\alpha + \beta = 1$ and that $\alpha' + \beta' = 1$, although these assumptions are not needed in the most general statement of the generic utility theory. The generic utility theory is a set of preference assumptions from which the existence of an interval scale satisfying (2) or (3) can be derived. The axiomatizations of (2) and (3) do not reduce in a simple way to previous utility axiomatizations (e.g., Debreu, 1959) because the coefficients, α and β, and α' and β', can differ on lower and upper triangular sets of gambles. Miyamoto (in press) showed that an interval scale satisfying (2) or (3) can be derived from assumptions that restrict standard additive conjoint measurement assumptions to the preference ordering on a triangular set of gambles.

Many well known utility theories imply that (2) and (3) are satisfied. EU theory implies (2) and (3) with $\alpha = \alpha' = p$ and $\beta = \beta' = 1 - p$ (Luce & Raiffa, 1957); SEU theory also implies (2) and (3), but it drops the requirement that the coefficients equal p and $1 - p$, respectively (Coombs, Dawes, & Tversky, 1970; Tversky, 1967). Edwards' ASEU and NASEU models imply that (2) and (3) are satisfied but they drop the requirement that $\alpha + \beta = 1$ and $\alpha' + \beta' = 1$ (Edwards, 1962). Karmarkar's (1978) SWU model implies that (2) and (3) are satisfied with $\alpha = \alpha' = p^{2\theta}/(p^{2\theta} + q^{2\theta})$ and $\beta = \beta' = q^{2\theta} + q^{2\theta})$, where $q = 1 - p$ and $\theta > 0$. The dual bilinear model implies that (2) is satisfied by preferences between gambles in any lower triangular set and that (3) is satisfied by preferences between gambles in any upper triangular set, but unlike the preceding theories it leaves open the possibility that $\alpha \neq \alpha'$ and $\beta \neq \beta'$ (Luce & Narens, 1985; Narens & Luce, 1986). Prospect theory also implies that (2) and (3) are satisfied by certain subsets of p-gambles, while leaving open the possibility that $\alpha \neq \alpha'$ and that $\beta \neq \beta'$. We will not attempt to state prospect theory and the dual bilinear model here, nor will we prove that they imply (2) and (3). (For statements of these theories, see Kahneman & Tversky, 1979; Luce & Narens, 1985; Narens & Luce, 1986; for a demonstration that they imply (2) and (3), see Miyamoto, 1987a, 1988). It is necessary to describe certain features of prospect theory, however, because they are pertinent to the interpretation of the experiment reported below. Our discussion omits many important aspects of prospect theory, focusing only on aspects that are needed to interpret subsequent theoretical and experimental work.

According to prospect theory, outcomes are perceived as gains or losses relative to a neutral reference level, which we denote as s_0. An *irregular p-gamble* is a p-gamble whose outcomes are both nonlosses or both nongains,[2] i.e., a gamble $(x\ p\ y)$ such that $x \geqslant_p s_0$ and $y \geqslant_p s_0$, or such that $s_0 \geqslant_p x$ and $s_0 \geqslant_p y$. Let G_1 be the set of $(x\ p\ y)$ such that $x \leqslant_p y \leqslant_p s_0$, and let G_2 be the set of $(x\ p\ y)$ such that $s_0 \leqslant_p x \leqslant_p y$. Note that G_1 and G_2 are both upper triangular sets of irregular p-gambles. It is easy to show that prospect theory implies that (3) is satisfied by the gambles in G_1 and also by the gambles in G_2 (Miyamoto, 1988).[3] Suppose that the stimulus gambles in an experiment have all been chosen to have the form $(x\ p\ y)$ where $x \leqslant_p y$. If the reference level of a subject is sufficiently high, then $x \leqslant_p y \leqslant_p s_0$ for every $(x\ p\ y)$ in the stimulus set. For such a subject, every stimulus gamble would be in G_1. Similarly, if the reference level of a subject is sufficiently low, then $s_0 \leqslant_p x \leqslant_p y$ for every $(x\ p\ y)$ in the stimulus set. For such a subject, every stimulus gamble would be in G_2. Now suppose that one samples enough subjects so that one finds a reasonably large subsample of individuals for whom the stimulus gambles are exclusively in G_1 or exclusively in G_2. When tests that are formulated in the generic utility framework are applied to the data from such subjects, the results of the test are interpretable from the standpoint of prospect theory, because prospect theory implies that they evaluate the stimulus gambles by a rule of the form (3).

In this section, we have argued that a number of major theories of preference under risk imply that the generic utility representation is satisfied by appropriately chosen subsets of gambles and subjects. Next, we will define the log/power and linear/exponential utility models and show how to formalize these models in the generic utility framework. Tests of these models in the generic utility framework will be interpretable from the standpoint of all theories that imply that the generic utility representation is satisfied.

The Log/Power and Linear/Exponential Families of Utility Models

An increasing utility function will be said to be a member of the log/power family if it has one of the forms

$$U(x) = \rho x^{\omega} + \tau, \quad \text{where } \rho > 0 \text{ and } \omega > 0, \tag{4}$$
$$U(x) = \rho(\log x) + \tau, \quad \text{where } \rho > 0, \tag{5}$$
$$U(x) = \rho x^{\omega} + \tau, \quad \text{where } \rho < 0 \text{ and } \omega < 0. \tag{6}$$

Risk posture is risk seeking if $\omega > 1$, risk neutral if $\omega = 1$, and risk averse if $\omega < 1$. The logarithmic utility function, Eq. (5), is more risk averse than any of the positive power utility functions, and the negative power utility functions, Eq. (6), are more risk averse than the logarithmic utility function. The logarithmic function is grouped together with the power utility functions because it fills the gap in the power functions created by the case where $\omega = 0$. As ω approaches zero from either the positive or negative direction, the preference ordering over gambles determined by a power utility function becomes increasingly like the preference ordering of the logarithmic function. In this sense, the logarithmic utility function is the limiting case of a power utility function as ω approaches zero. In the log/power family of utility functions, risk posture is completely determined by the value of ω, where it is understood that U is logarithmic if $\omega = 0$ (Keeney & Raiffa, 1976). As we will see, the log/power utility functions are characterized axiomatically by a simple property of preference behavior.

An increasing utility function will be said to be a member of the linear/exponential family if it has one of forms

[2] The definition of irregular gamble stated here is slightly different from that of Kahneman and Tversky (1979), but it is consistent with their terminology. See Miyamoto (1987a, in press) for an analysis of the distinction between regular and irregular gambles in prospect theory.

[3] Prospect theory also implies that other classes of gambles satisfy (2) and (3), but these gambles are not relevant to the present experiment. See Miyamoto (1987a, in press).

$$U(x) = \lambda e^{\theta x} + \tau \quad \text{for some } \lambda, \theta > 0, \tag{7}$$
$$U(x) = \lambda x + \tau \quad \text{for some } \lambda > 0 \text{ and some } \tau, \tag{8}$$
$$U(x) = \lambda e^{\theta x} + \tau \quad \text{for some } \lambda, \theta < 0. \tag{9}$$

In the linear/exponential family, an increasing utility function is risk seeking if Eq. (7) holds, it is risk neutral (linear) if Eq. (8) holds, and it is risk averse if Eq. (9) holds. As θ approaches zero from either the positive or negative direction, the preference ordering over gambles that is implied by an exponential utility function becomes increasingly like the preference ordering that is implied by a linear function. The linear utility function fills the gap in the exponential utility functions created by the case where $\theta = 0$. In the linear/exponential family of utility functions, risk posture is completely determined by the value of θ, where it is understood that U is linear if $\theta = 0$ (Keeney & Raiffa, 1976). The exponential parameter θ provides a summary measure of risk posture within the linear/exponential utility functions, just as the power ω summarizes risk posture within the log/power utility functions. The linear/exponential utility functions are also characterized axiomatically by a simple property of preference behavior.

Pratt (1964) formulated the standard axiomatizations of the log/power and linear/exponential utility functions (see also Keeney & Raiffa, 1976). For any gamble g and number t, let $t * g$ denote the gamble that is like g except that the outcomes of g have been multiplied by t, and let $g + t$ denote the gamble that is like g except that t has been added to the outcomes of g. (We are assuming that the outcomes are quantified objects like money or survival duration, so that $t * g$ and $g + t$ are meaningful operations.) Under the assumptions that EU theory is valid, that the utility function is twice differentiable, and that the limit of $-xU''(x)/U'(x)$ exists as x approaches 0, Pratt proved that the utility function is a logarithmic or power function provided that

$$g \geqslant_{\mathrm{p}} g' \quad \text{iff} \quad t * g \geqslant_{\mathrm{p}} t * g' \tag{10}$$

for every g and g', and positive number t. Furthermore, he proved that the utility function must be linear or exponential if

$$g \geqslant_{\mathrm{p}} g' \quad \text{iff} \quad g + t \geqslant_{\mathrm{p}} g' + t \tag{11}$$

for every g and g', and any number t. Pratt's formalization is seriously limited by the assumption that EU theory is valid. Psychological research has demonstrated that preference judgments systematically violate EU and SEU theory (Grether & Plott, 1979; Kahneman & Tversky, 1979; Slovic & Lichtenstein, 1983; Von Winterfeldt & Edwards, 1986).

Miyamoto (in press) pointed out that Pratt's formalization could be reformulated under the weaker assumptions of the generic utility theory. This reformulation will be stated here in terms of an upper triangular set of p-gambles, although a lower triangular set of p-gambles would serve equally well in the formalization.

Let G denote an upper triangular set of p-gambles. If the generic utility representation is satisfied, and if the utility scale is a continuous, increasing function of duration, then the following condition is both necessary and sufficient for U to be a power or logarithmic function: for any $(w\ p\ x)$, $(y\ p\ z)$ in G and any positive number s,

$$(w\ p\ x) \geqslant_{\mathrm{p}} (y\ p\ z) \quad \text{iff } (sw\ p\ sx) \geqslant_{\mathrm{p}} (sy\ p\ sz). \tag{12}$$

Furthermore, the following condition is both necessary and sufficient for U to be a linear or exponential function: for any $(w\ p\ x)$, $(y\ p\ z)$ in G and any positive number t,

$$(w\ p\ x) \geqslant_{\mathrm{p}} (y\ p\ z) \quad \text{iff } (w + t\ p\ x + t). \tag{13}$$

The proofs that (12) implies the log/power utility model and that (13) implies the linear/exponential utility models are derived from functional equations that were solved by Aczel (1966), Luce (1959), and Pfanzagl (1959). Miyamoto (in press) adapted their formalizations of the log/power and linear/exponential models to the generic utility framework. Conditions (12) and (13) are Pratt's (1964) conditions for the log/power and linear/exponential models, restricted to an upper triangular set of p-gambles. The present formalization generalizes Pratt's formalization in two ways. First, the strong differentiability assumptions in Pratt's formalization are replaced by the weaker assumption that the utility function is increasing and continuous. Second, the assumption that expected utility theory holds is replaced by the weaker assumption that there exists a generic utility representation for the preference ordering on an upper triangular set of p-gambles.[4] Because these assumptions are weak, experimental tests of (12) and (13) are interpretable from the standpoint of nonstandard utility theories as well as standard utility theories.

Certainty Matching Judgments and Proportional Matches

To describe how (12) and (13) are tested experimentally, we must first define the certainty matching judgments on which the experiment is based. The stimulus gambles used in our experiment were exclusively even-chance gambles, i.e., $p = .5$. Therefore we will describe the tests of

[4] The relations described here would also be satisfied if the upper triangular set of gambles were placed by a lower triangular set of gambles.

hypotheses in terms of even-chance gambles, but the tests would also be valid if some other value of p had been used.

Let $(x\ .5\ y)$ denote an even-chance gamble between x and y years of survival in good health. The certainty match of $(x\ .5\ y)$ will be defined to be a duration z that is judged to be subjectively equal in value to $(x\ .5\ y)$. In utility analysis, what we call a certainty match is more commonly known as the certainty equivalent of a gamble (Keeney & Raiffa, 1976; McNeil *et al.*, 1978; Von Winterfeldt & Edwards, 1986). We prefer to call it a certainty match because the terminology reminds us that in this method, the certain outcome is the dependent variable which has been chosen to match a gamble in subjective value. Let $M(x\ .5\ y)$ denote the certainty match of $(x\ .5\ y)$. By definition,

$$M(x\ .5\ y) \sim_p (x\ .5\ y). \tag{14}$$

It is easy to show that Eq. (12) holds if and only if

$$M(sx\ .5\ sy) = sM(x\ .5\ y) \tag{15}$$

for any x, y, and s that is not too large. (We assume that x, y, and s are confined to ranges that yield realistically reasonable survival durations, i.e., not negative durations or unreasonably long durations.) Similarly, it can be demonstrated that Eq. (13) holds if and only if

$$M(x + t\ .5\ y + t) = M(x\ .5\ y) + t \tag{16}$$

for any x, y, and t that is not too large.

When testing Eqs. (15) and (16), it is useful to transform the data from certainty matching judgments to another quantity, which we call proportional matches. Let g be an even-chance gamble, let L_g and H_g denote the lower and higher outcomes of g, and let $M(g)$ denote the judged certainty match of g on some trial. The transformation from a certainty match, $M(g)$, to a proportional match, $PM(g)$, is defined by the relation

$$PM(g) = \frac{M(g) - L_g}{H_g - L_g} \tag{17}$$

By definition, the proportional match, $PM(g)$, represents the proportion of the range of g that is exceeded by $M(g)$. It is not hard to show that Eq. (15) is equivalent to the following equation between proportional matches: for any x, y, and s,

$$PM(sx\ .5\ sy) = PM(x\ .5\ y). \tag{18}$$

Similarly, Eq. (16) is equivalent to the following equation between proportional matches: for any x, y, and t,

$$PM(x + t\ .5\ y + t) = PM(x\ .5\ y). \tag{19}$$

The proofs that (15) and (18) are equivalent and that (16) and (19) are equivalent are given in the Appendix. In the experiment reported below, a set of stimulus gambles was designed to test Eqs. (15) and (16). Certainty matches were elicited for these gambles, the data were transformed to proportional matches, and (18) and (19) were tested in any analysis of variance applied to the transformed data. The equivalence of (12), (15), and (18) establishes that the experiment tests the log/power utility models, and the equivalence of (13), (16), and (19) establishes that it tests the linear/exponential utility models.

There are three reasons for preferring to analyze proportional matches rather than certainty matches. First, Eqs. (18) and (19) have natural interpretations in an analysis of variance. The equations assert that mean proportional matches of gambles are unchanged by multiplication by a constant or addition of a constant, respectively. Second, the variance of certainty matching judgments increases with the range of the stimulus gamble. Transforming certainty matching data to proportional matches tends to equalize the variance of responses to different gambles. Third, sample means are more directly interpretable when expressed as proportional matches. If we find that the mean proportional match was .4 in some condition, we know that the average certainty match was slightly below the expected value of the gamble. If we find that the mean certainty match was 9.3 years in that condition, the mean cannot be interpreted until we determine the specific gambles used in the condition.

METHOD

Subjects

Subjects were 44 undergraduates at the University of Michigan for whom participation in a psychology experiment fulfilled a course requirement. Five subjects did not complete the experimental task, either because they worked too slowly to finish in the allocated time (3 hr) because they could not be taught the task of judging certainty matches, or because they were negligent in attending experimental sessions. In addition, the responses of one subject indicated that he had misunderstood the task. Therefore 6 subjects were dropped from the sample, and the effective sample size was 38 subjects.

Design

Table 1 lists the stimulus gambles for the experiment. There were six sets of gambles, Basic, Plus 10, Plus 20, Times 2, Times 3, and Zero. The Basic gambles consisted of all combinations of 1, 2, 3, and 4 years with 10 or 12 years of survival. The Plus 10 and Plus 20 gambles were produced by adding 10 years and 20 years, respectively, to the outcomes of the

TABLE 1
STIMULUS GAMBLES[a]

Basic					Zero
1/10	2/20	3/30	11/20	21/30	0/32
2/10	4/20	6/30	12/20	22/30	0/36
3/10	6/20	9/30	13/20	23/30	
4/10	8/20	12/30	14/20	24/30	
1/12	2/24	3/36	11/22	21/32	
2/12	4/24	6/36	12/22	22/32	
3/12	6/24	9/36	13/22	23/32	
4/12	8/24	12/36	14/22	24/32	

[a] X/Y denotes the outcomes of an even-change gamble between X and Y years of survival

Basic gambles. The Times 2 and Times 3 gambles were produced by multiplying the outcomes of the Basic gambles by 2 and 3, respectively. The Zero gambles consisted of the two gambles, (0 .5 32) and (0 .5 36). In every stimulus gamble, the shorter outcome preceded the longer outcome. Thus every gamble had the form of an upper triangular gamble. Gambles were presented to subjects in four blocks of trials. Within a block, each gamble appeared once, with different random orders of presentation in the different blocks. Each block was presented to a subject as a written questionnaire. Individual items consisted of a gamble for survival duration together with a query as to the certain survival that matched the gamble in preference.

Leaving aside the data for Zero gambles, the data for an individual subject can be represented as an 8×5 two-way fixed-effects ANOVA. The 8 levels of the first factor are the 8 Basic gambles. The 5 levels of the second factor are the 5 transformations (null, Plus 10, Plus 20, Times 2, and Times 3) that are applied to the Basic gambles. This design will be referred to as the Gamble $\times$ Transformation (G $\times$ T) design. It has 40 cells and 4 replications per cell. The data for the Basic, Times 2, and Times 3 gambles provide tests of conditions (12), (15), and (18). The data for the Basic, Plus 10, Plus 20 gambles provide tests of conditions (13), (16), and (19). It is also possible to test specifically whether a utility function is logarithmic or a negative power. A test of these hypotheses will be described more fully in the results section.

Procedure

Subjects were tested individually during two experimental sessions on separate days. At the first session, the experimenter explained the task of judging certainty matches using practice gambles as examples. The subject was asked to judge the certainty matches of practice gambles, while the experimenter would comment on questions that the subject would

raise about the task. Subject were instructed to bear the following conditions in mind while judging certainty matches. First, both outcomes of the gamble should be regarded as accompanied by good health. Similarly, the certain outcome matched to the gamble should also be assumed to be in good health. The precise definition of good health was left to the subject. Second, any nonzero survival duration should be interpreted as a survival lasting that many years, but no more than that many years. For example, an 8-year survival signified a survival lasting a full 8 years, but no more than 8 years. The hypothetical cause of death was left unspecified, but the subject was instructed to assume that death would be sudden and relatively painless. Third, the subject was instructed to assume that any survival would be lived in the normal state of ignorance or incomplete knowledge as to the date of one's death. This last instruction was included to prevent subjects from inferring that the certain survival had the added value of allowing one to know how much time one had to live.

Subjects were taught to bracket the matching duration prior to responding with a certainty match. For example, when attempting to match a gamble between 2 and 24 years of survival, the subject should initially note, e.g., that 4 years is too low to be the match, and that 22 years is too high to be a match. The match should be found by successively considering what outcomes would be too low or too high to be the match. Subjects were instructed to bracket the match before producing it in order to avoid anchoring and adjustment biases that can be present in subjective judgment (Tversky & Kahneman, 1974). The subject and experimenter continued to discuss the task until matches had been produced for nine practice gambles. If the experimenter was not confident that the subject correctly understood the task, they would work through the certainty matches of additional practice gambles until the subject appeared to understand the task.

Once instruction in certainty matching was completed, the subject's reference level of survival duration was assessed. We attempted to identify a subject's reference level by defining the concept for him, and asking him to judge his own reference level. In discussing the reference level with the subject, the experimenter referred to it as the "aspiration level of survival." The concept was explained as follows:

> Next, I'm going to ask you about something called the "aspiration level of survival." Since this concept is fairly complicated, I'll explain it in several steps. The aspiration level for survival is defined to be the length of survival that marks the boundary between those survivals that you regard as a loss and those survivals that you regard as a gain.
>
> For example, my own aspiration level for survival is about the age of sixty. This means that if I found out that I were going to live to the age of fifty or fifty five (but no more), I would regard this as something of a loss. If I found out that I were going to live to the age of sixty five or seventy, I would regard this as something of a gain.

> The aspiration level is not the same as my life expectancy, since my life expectancy is greater than sixty. It's not the length of time that I would want to live, since I'm sure I would want to live longer if I were in good health.
>
> I should mention that there's nothing special about the age of sixty. Some individuals place their aspiration level at a large number, like ninety. For such a person any survival less than ninety would be regarded to some degree as a loss. I've also encountered individuals who set their aspiration level at their present age. This does not mean that they no longer want to live. It means that they regard every year of survival as a gain. If such an individual learned that he had two years to live, he would regard this as gaining two years of survival, rather than to emphasize some longer survival of which he is being deprived.

Subjects generally found this explanation meaningful, and would state a value for the reference level without appearing to be confused or uncertain. If the subject needed further explanation of the concept, the experimenter would ask whether specific durations were perceived as losses or gains. Would you regard 2 years of survival as a loss or a gain? Would you regard 60 years of survival as a loss or a gain? Subjects found these questions to be meaningful. Having recognized that survivals could be classified as losses or gains, it became possible to identify an age that marked the boundary between losses and gains.

Once elicitation of the reference level was completed, the subject was given a written questionnaire which contained the stimulus gambles to be judged for certainty matches. Four different questionnaires were prepared, each questionnaire listing the 42 stimulus gambles in a different random order. At the first experimental session, the subject completed the responses to the first two questionnaires. At the second session, the subject completed the second two questionnaires. A subject never had different questionnaires in his possession at the same time. Thus comparison of responses across questionnaires was impossible. Furthermore, subjects were instructed to state a certainty match for each gamble on a questionnaire without consideration for matches assigned to previous gambles on the questionnaire.

PRELIMINARY ANALYSES

Before describing the main results, it is necessary to describe the calculation of a within-subject mean-squared error, and tests of the consistency of certainty matching with the additive representation postulated in Eq. (3).

Within-Subject MSE and Violations of Betweenness

A separate mean-squared error (MSE) was calculated for each subject using the data in the 40 cells of the G × T design. To explain precisely what calculation was made, we need to describe violations of a property called betweenness.

Recall that every stimulus gamble had the form $(x\ .5\ y)$ where $x < y$. A certainty matching judgment $M(x\ .5\ y)$ will be said to satisfy *betweenness* iff $x \leqslant M(x\ .5\ y) \leqslant y$. Betweenness is a compelling principle of rational choice for one would surely prefer $(x\ .5\ y)$ to its lower outcome x, but not to its upper outcome y. Coombs and Huang (1976) found systematic violations of a more general betweenness property in an experiment with complex, four-outcome gambles. The stimulus gambles in the present experiment were especially simple, however, and subjects readily recognized that certainty matches should satisfy betweenness. Therefore we attribute violations of betweenness to performance errors such as misreading the gamble or a careless response. (A large number of violations by a single subject would suggest that the subject did not understand the task, but this did not occur.) To reduce the influence of matching judgments that violated betweenness, such matches were reassigned the closest value within the range of the gamble. In other words, if $M(x\ .5\ y) < x$, then $M(x\ .5\ y)$ was reassigned the value x. If $M(x\ .5\ y) > y$, then $M(x\ .5\ y)$ was reassigned the value y. The certainty matches were then transformed to proportional matches as in Eq. (17). If the original value of $M(x\ .5\ y)$ violated betweenness, then the corresponding proportional match was either 0 or 1, depending on whether the certainty match was reassigned the value x or y. An MSE for a given subject was calculated in the standard way from the proportional matches of the G × T ANOVA. This MSE had 120 $(= 8 \times 5 \times 3)$ degrees of freedom.

There was no evidence for systematic violations of betweenness in the experimental data. A subject had 160 opportunities to violate betweenness. Twenty-four subjects had 0 violations, 9 had 1 violation, and 2 had 2 violations. The 3 remaining subjects had 3, 4, and 6 violations, respectively. We note that these subjects had the seventh, first, and second largest MSE, where the MSE was calculated after removal of violations of betweenness. Thus, violations of betweenness were positively related to variability of response, making it plausible that the violations were due to careless performance

Violations of Dominance

A gamble $(w\ .5\ x)$ is said to dominate a gamble $(y\ .5\ z)$ if $w > y$ and $x \geqslant z$, or else, $w \geqslant y$ and $x > z$. It is reasonable to suppose that if g dominates g', then $g >_p g'$. We will say that dominance has been violated if g dominates g', but $g' \geqslant_p g$. We checked for violations of dominance as follows. A mean certainty match, denoted $\overline{M}(w\ .5\ x)$, was calculated within subject for each gamble of the form $(w\ .5\ x)$. A violation of dominance was recorded each time $(w\ .5\ x)$ dominated $(y\ .5\ z)$ and $\overline{M}(y\ .5\ z) \geqslant$

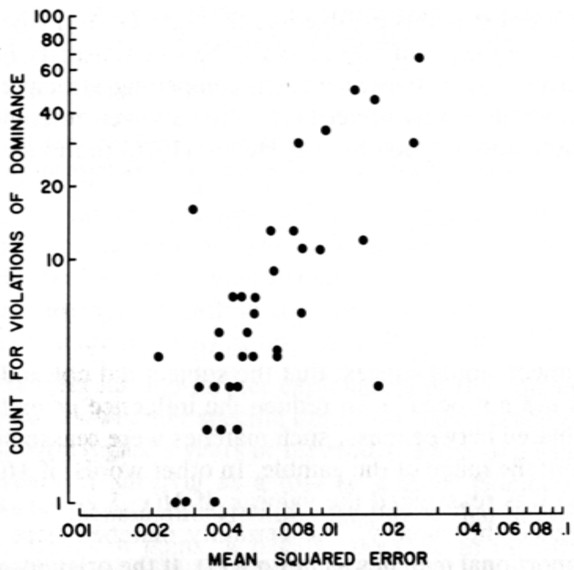

FIG. 1. Individual subject mean-squared error (MSE) versus counts for violations of dominance on logarithmic coordinates. The one observation not shown in the figure had an MSE of .0066 and a count of 0.

$\overline{M}(w\ .5\ x)$.[5] The total number of violations was counted for each subject. It can be shown that there are 602 possible violations of dominance among the 40 gambles in the G × T factorial design. Figure 1 shows MSE plotted against violations of dominance on logarithmic coordinates. It is clear that violations of dominance were positively related to MSE, as one would expect if violations of dominance were due to response variation rather than some systematic tendency.

Test of Additivity

The generic utility representation asserts that the utility of a gamble is an additive function of the utilities of its outcomes. To test whether the matching judgments of individual subjects were consistent with additivity, an ordering of the stimulus gambles was inferred for each subject from the order of the subject's mean certainty matches. In other words, $(w\ .5\ x) \geq_p (y\ .5\ z)$ was inferred if $\overline{M}(w\ .5\ x) \geq \overline{M}(y\ .5\ z)$. A computer program called ORDMET (McClelland & Coombs, 1975) was used to determine the extent to which the inferred ordering of gambles was consistent with an additive representation of the form

$$(w\ .5\ x) \geq_p (y\ .5\ z) \quad \text{iff } \phi(w) + \psi(x) \geq \phi(y) + \psi(z). \tag{20}$$

The ORDMET program tests whether a set of inequalities is consistent with the existence of functions ϕ and ψ satisfying (20). If the inequalities are inconsistent with (20), ORDMET reduces the set of inequalities until a subset is found that is consistent with (20). It should be noted that (20) is not equivalent to (3) because (3) makes the additional claim that ϕ is linear with respect to ψ. In effect, ORDMET tests whether (20) can be satisfied by the data, but does not check whether ϕ and ψ can be chosen to be linear with respect to each other. Because of hardware limitations on computer memory, only the preferences between the 32 gambles in the Basic, Times 2, and Times 3 conditions were used in the test of (20). The input to ORDMET was further reduced by omitting any inequality that violated dominance. Violations of dominance were omitted because the preceding analysis already provides a count of violations of dominance.

It was found that the $\geq_p$orderings of 26 subjects were completely consistent with (20). The orderings of 11 subjects were consistent with (20) after eliminating one inequality from the subject's data, and the ordering of one other subject was consistent with (20) after eliminating two inequalities. Therefore, excluding violations of dominance, the orderings of certainty matches for the Basic, Times 2, and Times 3 gambles were quite consistent with the additive representation.

[5] Technically, a violation of betweenness implies a violation of dominance because $M(x\ .5\ y) \leq x$ implies $(x\ .5\ y) \leq_p x = (x\ .5\ x)$ and $M(x\ .5\ y) \geq y$ implies $(x\ .5\ y) \geq_p y = (y\ .5\ y)$. These violations of dominance have already been discussed. The analysis of this section pertains to violations of dominance in preferences between gambles rather than between gambles and certainty matches.

MAIN RESULTS

Reference Level for Survival Duration

Let A denote a subject's present age, and let s_0 denote the reference level of survival duration judged by the subject. The outcomes of the stimulus gambles ranged from 0 to 36 years. Therefore, in the terminology of prospect theory, every stimulus gamble was irregular for any subject who judged $s_0 = A$ or $s_0 \geq A + 36$.[6] Of the 38 subjects, 4 judged $s_0 = A$, 7 judged $A < s_0 < A + 36$, and 27 judged $A + 36 \leq s_0$. The fact that most subjects judged $s_0 \geq A + 36$ is not surprising in view of the fact that the ages of subjects ranged from 18 to 28 years, with a mean of 20.3 years. Although the judgment by some subjects that $s_0 = A$ may seem counterintuitive, this judgment expresses a desire to view each year of survival as a gain. For example, one subject who regarded s_0 equal to A related how

[6] See footnote 2.

his wife had almost died from head injuries in a horse riding accident. The experience had brought home to him how each day of life was a blessing. Subjects who judged $s_0 = A$ or $s_0 \geqslant A + 36$ will be referred to as external reference level (ERL) subjects; their reference levels are on the boundary or external to the range of outcomes in the stimulus gambles. Subjects who judged $A < s_0 < A + 36$ will be called internal reference level (IRL) subjects.

Because the integration rule in prospect theory varies depending on the magnitude of gamble outcomes relative to the reference level, prospect theory predicts that the certainty matching judgments of ERL subjects are determined by a single integration rule (every stimulus gamble was irregular for these subjects), but the certainty matching judgments of IRL subjects could be determined by different integration rules depending on whether a stimulus gamble happened to be regular or irregular relative to the subject's reference level. The tests of parametric models presented here are valid only if a single integration rule is satisfied by a subject's matching judgments for every stimulus gamble. Hence, only the results for ERL subjects can be interpreted from the standpoint of prospect theory. EU and SEU theory, the SWU model, and the dual bilinear utility model postulate that the matching judgments of all subjects satisfy a generic utility representation. Therefore the results for all subjects can be interpreted from the standpoint of these theories. We will first present the results for all subjects combined, and then restrict the analysis to the ERL subjects.

Linear/Exponential and Log/Power Utility Models: All Subjects

Let $(x\ .5\ y)$ be any Basic gamble (see Table 1). The linear/exponential model implies that

$$PM(x\ .5\ y) = PM(x + 10\ .5\ y + 10) = PM(x + 20\ .5\ y + 20) \qquad (21)$$

Because matching judgments are subject to random variation, (21) should be regarded as a property of the underlying population means of matching judgments. Interpreted in this way, (21) asserts that the gambles in (21) satisfy the null hypothesis of a one-way analysis of variance (ANOVA). Furthermore, as (21) is predicted to hold of every Basic gamble in the stimulus set, it can be tested in a two-way ANOVA that is defined as follows. Let G denote the 8 levels of the Gambles factor, and let T^+ denote the Basic, Plus 10, Plus 20 levels of the Transformation factor (see Table 1). Equation (21) is satisfied by every Basic gamble if and only if there is no main effect of T^+, and no interaction between G and T^+. Hence, (21) can be rejected if the G $\times$ T^+ ANOVA yields a significant main effect of T^+ or a significant G $\times$ T^+ interaction. A G $\times$ T^+ ANOVA

was calculated for each subject using the previously described MSE as the estimate of error. The sums of squares for the main effect of T^+ and for the G $\times$ T^+ interaction were calculated in the usual way. The F for T^+ had 2 and 120 degrees of freedom, and the F for the G $\times$ T^+ interaction had 14 and 120 degrees of freedom. The main effect of T^+ was significant at the .05 level for 22 subjects (58%) and the G $\times$ T^+ interaction was significant at the .05 level for 26 subjects (68%). Only 7 subjects (18%) were nonsignificant for both T^+ and G $\times$ T^+. Clearly, the linear/exponential utility model can be rejected for the majority of subjects.

A similar analysis can be made for the log/power utility model. This model implies that

$$PM(x\ .5\ y) = PM(2x\ .5\ 2y) = PM(3x\ .5\ 3y) \qquad (22)$$

Let T^* denote the Basic, Times 2, and Times 3 levels of the Transformation factor, A G $\times$ T^* ANOVA analogous to the G $\times$ T^+ ANOVA was calculated for each subject. Equation (22) can be rejected if the G $\times$ T^* ANOVA yields a significant main effect of T^* or a significant G $\times$ T^* interaction. The main effect of T^* was significant for 29 subjects (76%) and the interaction of G and T^* was significant for 24 subjects (63%). Only 4 subjects (11%) were nonsignificant for both T^* and G $\times$ T^*. It is clear that the log/power utility model can be rejected for the majority of subjects.

The results show clearly that the linear/exponential and log/power utility models were frequently violated, but it might be argued that the criterion for rejecting the models was excessively liberal. For example, we rejected the linear/exponential model if either the main effect of T^+ or the G $\times$ T^+ interaction was significant. If two null hypotheses are tested at the .05 level, the probability of rejecting at least one hypothesis when both are true is between .05 and .10 (by the Dunn–Bonferroni multiple comparison procedure). Thus, the preceding analysis tests the linear/exponential and log/power utility models at a level of significance between .05 and .10. To answer this objection, we calculated an F statistic that combines the tests for the main effect and interaction effect. In the case of the linear/exponential model, we pooled the sums of squares for the main effect of T^+ and the interaction of G and T^+, and divided this sum by the pooled degrees of freedom (=16). Dividing this quantity by the MSE yields a statistic that is distributed as $F(16,120)$ when there is no main effect of T^+ and no G $\times$ T^+ interaction. Similarly, for the log/power model, the sums of squares for the main effect of T^* and the interaction of G and T^* were pooled, divided by the pooled degrees of freedom, and divided by the MSE to yield an F test for the hypothesis that the main effect of T^* and the G $\times$ T^+ interaction were both zero. The results of this analysis are displayed in Table 2. The linear/exponential utility model was rejected at the .05 level by 28 subjects (74% ± 14% with 95% confidence). The log/power utility model was rejected at the .05 level by 35 subjects (92% ± 9% with 95% confidence). Furthermore, both classes of models were rejected by 28 subjects (74% ± 14% with 95% confidence). It is clear that a majority of individuals reject both classes of utility models.

Although these results give a quantitative estimate of the frequency with which these utility models are violated, the pattern of violations can be seen more easily in the scatter plots shown in Fig. 2. We will explain the symbols in the upper left scatter plot; the other scatter plots are analogous. Each point in the upper left scatter plot represents the mean proportional matches of an individual subject in the Basic and Times 2 conditions. The empty and filled circles represent ERL subjects; the empty and filled triangles represent IRL subjects. A t test was calculated within each subject's data for a mean difference between the Basic and Times 2 conditions. The error estimate used in the t test was the previously described MSE. The empty symbols in the scatter plot indicate that a subject's matches in the Basic and Times 2 conditions were significantly different at the .05 level. The filled symbols indicate that the proportional matches were not significantly different.

Consider the top three panels of Fig. 2 which show the pairwise comparisons of the Basic, Times 2, and Times 3 conditions. If the log/power utility models were precisely true of every subject, the points would be massed along the diagonals with deviations from the diagonals due only to random variation. Furthermore the expected frequency of significant t tests would be 5%, corresponding to the frequency of type I errors when the null hypothesis is true of every subject. Clearly, none of these scatter

TABLE 2

TEST OF THE LINEAR/EXPONENTIAL MODEL VERSUS TEST OF THE LOG/POWER MODEL FOR ALL SUBJECTS

		Log/power utility model		
		$p < .05$	$p > .05$	
Linear/exponential utility model	$p < .05$	28 74%	0 0%	28 74%
	$p > .05$	7 18%	3 8%	10 26%
		35 92%	3 8%	

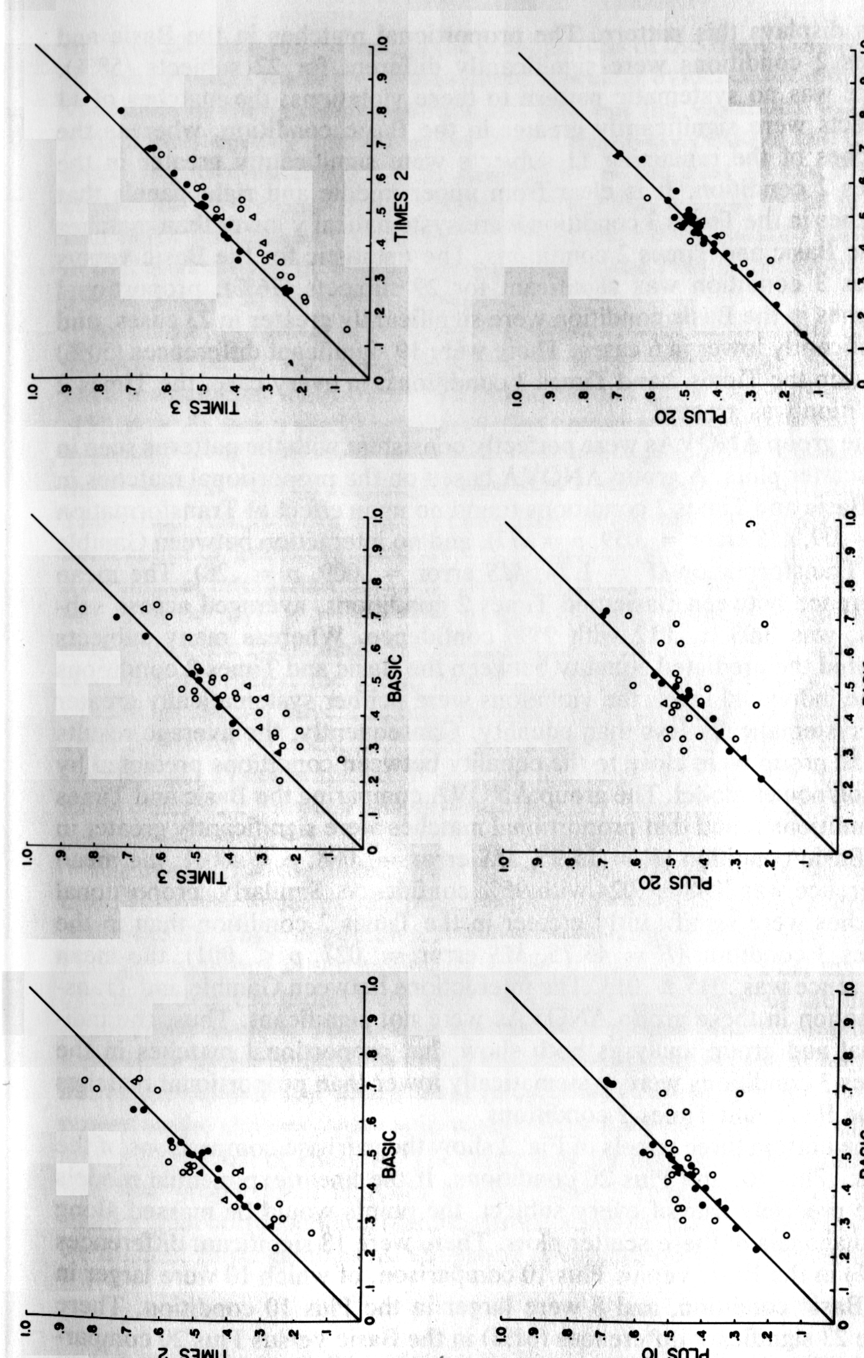

FIG. 2. Scatter plots of mean proportional matches. Each point represents the mean proportional matches in the pair of conditions that label the axes. Open circles and triangles stand for subjects with significant ($p < .05$) t statistics for a difference in means. Closed circles and triangles stand for subjects with nonsignificant differences. Solid and open circles stand for ERL subjects. Solid and open triangles stand for IRL subjects.

plots displays this pattern. The proportional matches in the Basic and Times 2 conditions were significantly different for 22 subjects (58%). There was no systematic pattern to these violations; the matches of 11 subjects were significantly greater in the Basic condition, whereas the matches of the remaining 11 subjects were significantly greater in the Times 2 condition. It is clear from upper middle and right panels that matches in the Times 3 condition were systematically lower than matches in the Basic and Times 2 conditions. The t statistic for the Basic versus Times 3 condition was significant for 29 subjects (76%); proportional matches in the Basic condition were significantly greater in 23 cases, and significantly lower in 6 cases. There were 19 significant differences (50%) between the Times 2 and Times 3 conditions; in every case, the Times 2 condition was greater.

The group ANOVAs were perfectly consistent with the patterns seen in the scatter plots. A group ANOVA based on the proportional matches in the Basic and Times 2 conditions found no main effect of Transformation ($F = .09$, MS error $= .059$, $p = .77$), and no interaction between Gamble and Transformation ($F = 1.24$, MS error $= .009$, $p = .28$). The mean difference between Basic and Times 2 conditions, averaged across subjects, was $.003 \pm .012$ with 95% confidence. Whereas many subjects violated the predicted equality between the Basic and Times 2 conditions at the individual level, the violations were neither systematically greater nor systematically less than equality. Consequently, the average results for the group were close to the equality between conditions predicted by the log/power model. The group ANOVA comparing the Basic and Times 3 conditions found that proportional matches were significantly greater in the Basic condition ($F = 15.95$, MS error $= .088$, $p < .001$); the mean difference was $.048 \pm .024$ with 95% confidence. Similarly, proportional matches were significantly greater in the Times 2 condition than in the Times 3 condition ($F = 45.75$, MS error $= .027$, $p < .001$); the mean difference was $.045 \pm .013$. The interactions between Gamble and Transformation in these group ANOVAs were not significant. Thus, the individual and group analyses both show that proportional matches in the Times 3 conditions were systematically lower than proportional matches in the Basic and Times 2 conditions.

The bottom three panels of Fig. 2 show the pairwise comparisons of the Basic, Plus 10, and Plus 20 conditions. If the linear/exponential models were precisely true of every subject, the points would be massed along the diagonals of these scatter plots. There were 18 significant differences (47%) in the Basic versus Plus 10 comparison, of which 10 were larger in the Basic condition, and 8 were larger in the Plus 10 condition. There were 23 significant differences (61%) in the Basic versus Plus 20 comparison, of which 12 were larger in the Basic condition and 11 were larger in

the Plus 20 condition.[7] Thus, in the comparisons of the Basic condition to the Plus 10 and Plus 20 conditions, there was clear evidence for individual violations of the linear/exponential models, but the violations were approximately equally distributed between excessively positive and excessively negative differences. The scatter plot of the Plus 10 and Plus 20 conditions (lower right panel) shows a pattern that is close to the prediction of the linear/exponential models. There were 8 significant t statistics in this comparison, of which 5 were greater in the Plus 10 condition, and 3 were greater in the Plus 20 conditions.

As before, the group ANOVAs are consistent with the patterns seen in the scatter plots. The group ANOVA for Basic versus Plus 10 did not have a significant main effect of Transformation ($F = .42$, MS error $= .151$, $p > .5$), nor a significant interaction between Gamble and Transformation ($F = .42$, MS error $= .011$, $p > .5$). The mean difference between the Basic and Plus 10 conditions was $.010 \pm .032$ with 95% confidence. The group ANOVA for Basic versus Plus 20 did not have a significant main effect of Transformation ($F = 1.17$, MS error $= .207$, $p > .25$), nor a significant interaction between Gamble and Transformation ($F = 1.12$, MS error $= .011$, $p > .25$). The mean difference between proportional matches in the Basic and Plus 20 conditions was $.020 \pm .037$ with 95% confidence. Finally, the group ANOVA comparing Plus 10 and Plus 20 conditions did not have a significant main effect of Transformation ($F = 2.93$, MS error $= .020$, $p > .05$) nor a significant interaction between Gamble and Transformation ($F = 1.88$, MS error $= .010$, $p > .05$). The mean difference between proportional matches in the Plus 10 and Plus 20 conditions was $.010 \pm .012$ with 95% confidence.

The results for the Basic, Plus 10, and Plus 20 conditions showed a close fit to the linear/exponential model at the level of group averages. Although the Basic condition was inconsistent with both the Plus 10 and Plus 20 conditions at the individual subjects' level, proportional matches averaged over the group were very close to the predictions of the linear/exponential models. The results for the Plus 10 and Plus 20 conditions were close to the predictions of the linear/exponential models at both the individual subject and group level of analysis. The results for the log/power model showed systematic violations at both the individual subject and group levels of analysis. The most striking feature of the violations was that proportional matches in the Times 3 condition tended to be smaller than corresponding matches in the Basic or Times 2 conditions. If we think of Basic gambles as comprised of short-term survival outcomes,

[7] One point with a significant t statistic in the Plus 10 versus Plus 20 comparison cannot be seen in Fig. 2 because it is overprinted by another point.

the Times 2 gambles as a mixture of short- and mid-term outcomes, and the Times 3 gambles as a mixture of short- and long-term outcomes, the results show that the utility of survival in the long-term does not increase as rapidly as would be predicted by a log/power model and the responses to gambles with short- or mid-term outcomes.

We also tested the hypotheses of a logarithmic or negative power utility function. Although these hypotheses are already disconfirmed by the preceding results, it is worth noting that a direct test of the hypothesis was embedded in the experimental design. If the utility function is either logarithmic or a negative power, the utility of 0 duration should be $-\infty$. Hence, an individual should prefer any certain survival however short to a gamble with 0 years as an outcome. The stimulus gambles included two Zero gambles, (0 .5 32) and (0 .5 36). The certainty match of each gamble was judged on 4 trials. Among all 38 subjects, the smallest mean certainty match for either Zero gamble was 3.3 years. Thus no subject produced an average certainty match that was even close to 0. For each subject, a *t* test was calculated for the hypothesis that the mean certainty match of Zero gambles was equal or less than 1.0 versus the hypothesis that it was greater than 1.0. We tested the hypothesis that the mean was less than or equal to 1.0 rather than that it was equal to 0.0 in order to show that the underlying mean of the certainty matches of Zero gambles could not even be close to 0.0. The hypothesis that the mean certainty match was less than or equal to 1.0 was rejected at the .05 level for all 38 subjects. The utility of 0 cannot be $-\infty$ as predicted by a logarithmic or negative power utility function.

Linear/Exponential and Log/Power Utility Models: ERL Subjects

The results for the 31 ERL subjects were very similar to the preceding results, so they will only be sketched. Table 3 presents the results of the pooled tests of the linear/exponential model and the log/power model for the ERL subjects alone, which were calculated in the same fashion as the results in Table 2. The linear/exponential model was rejected by 25 ERL subjects (81%), the log/power model was rejected by 30 ERL subjects (97%), and 25 ERL subjects (81%) rejected both models. Scatter plots of mean proportional matches in the various combinations of conditions were presented in Fig. 2. Empty and filled circles represent ERL subjects. It should be apparent that the pattern of results for the ERL subjects is qualitatively similar to the pattern for all subjects combined. Hence the previous discussion of these scatter plots with respect to all subjects applies without revision to the ERL subjects. Table 4 displays the differences in mean proportional matches for the pairs of conditions that test the log/power and linear/exponential models. Evidently, the pattern of results for ERL subjects was quite similar to the pattern for all subjects.

TABLE 3

TEST OF THE LINEAR/EXPONENTIAL MODEL VERSUS TEST OF THE LOG/POWER MODEL FOR EXTERNAL REFERENCE LEVEL SUBJECTS ONLY

		Log/power utility model		
		$p < .05$	$p > .0$[illegible]	
Linear/exponential utility model	$p < .05$	25 81%	0 0%	25 81%
	$p > .05$	5 16%	1 3%	6 19%
		30 97%	1 3%	

Since 31 of 38 subjects were ERL subjects, it is not surprising that the results for the ERL subjects were essentially identical to the results for all subjects. The analysis would be less redundant, however, if the results for all subjects had found that a parametric utility model was satisfied by all but a small minority of subjects. If this had been the case, one would have to check whether violations of predictions of the parametric utility model were due to invalidity of the generic utility representation rather than invalidity of the model. For example, if one assumes that the generic utility theory is true, then the log/power model implies the equality of mean proportional matches in the Basic, Times 2, and Times 3 conditions. Violations of this equality imply that either the log/power model or the generic utility representation is false. If prospect theory is valid, the judgments of ERL subjects must satisfy a generic utility representation relative to the stimulus gambles of the experiment, but it is possible that IRL subject violate the generic utility representation. Hence, if only the IRL subjects produced violations of the equality predicted by the log/power

TABLE 4

95% CONFIDENCE INTERVALS FOR DIFFERENCES BETWEEN LEVELS OF THE TRANSFORMATION VARIABLE

Comparison	All subject[illegible]	ERL subjects
Basic vs Plus 10	.010 ± .03[illegible]	.013 ± .039
Basic vs Plus 20	.020 ± .03[illegible]	.024 ± .046
Plus 10 vs Plus 20	.010 ± .01[illegible]	.010 ± .013
Basic vs Times 2	.003 ± .02[illegible]	.003 ± .023
Basic vs Times 3	.048 ± .02[illegible]	.046 ± .028
Times 2 vs Times	.045 ± .01[illegible]	.043 ± .015

model, one could attribute these violations to the invalidity of the generic utility representation, rather than to a violation of the log/power model. Similar remarks pertain to the linear/exponential model. In the present experiment, it was obvious that violations of the linear/exponential model and the log/power model and the log/power model were not restricted to IRL subjects.

Although the prospect theory analysis of results was not particularly informative in the present case, it is nevertheless important to see that the analysis exemplifies a methodology for testing utility models in the prospect theory framework. The application of this methodology depends on the identification of the reference levels of individual subjects. In the present study, we simply had subjects judge their own reference levels, but it is undoubtedly the case that the continued development of prospect theory will require further research on the methodology of identifying reference levels, and the factors that affect reference levels in specific judgment tasks (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981; Von Winterfeldt & Edwards, 1986). Determination of the reference levels of individuals is required in the present methodology in order to distinguish ERL from IRL subjects. If the stimulus gambles are all drawn from a lower triangular set or all from an upper triangular set, then prospect theory implies that ERL subjects, but not IRL subjects, satisfy the generic utility representation with respect to the stimulus gambles. Therefore classes of utility models that are formalized in the generic utility framework can be tested in the preference judgments of ERL subjects. The present study is devoted to tests of the log/power and linear/exponential utility models, but elsewhere, we have also tested a multiplicative utility model using this methodology (Miyamoto & Eraker, 1988). The generic utility theory makes possible the introduction of basic multiattribute utility models into prospect theory because these models can be formalized in the generic utility framework, and prospect theory implies that the generic utility representation must be satisfied by appropriately chosen subsets of gambles.

DISCUSSION

The present study has several methodological points to make, and it contributes substantively to the utility analysis of survival outcomes. We will discuss the methodological issues first.

Utility Analysis in the Generic Utility Framework

The generic utility representation provides a natural framework for utility analysis, especially if the primary concern is to investigate the

subjective structure of utility without attempting to address the many complex issues in the subjective representation of probability (Miyamoto, 1988). Although one would naturally prefer a valid theory that integrates utility and subjective probability, it is disadvantageous to develop utility analysis on a basis that is committed to a general theory of subjective probability when that theory is likely to undergo further revision. The generic utility theory is a provisional framework, and not an ultimate framework. If a strong theory of utility and subjective probability were found that commanded general acceptance among decision theorists, then one would naturally work in the framework of that theory. While we await the discovery of evidence and theoretical analysis that clearly support one theory over all competitors, we can nevertheless develop utility analysis in the generic utility framework with the assurance that conclusions will continue to be seen as valid from the standpoint of the many theories that imply it.

Of course, the generic utility theory is not an "assumptionless" theory, whatever that would mean. The measurement assumptions underlying the generic utility representation are essentially the additive conjoint measurement axioms of Krantz, Luce, Suppes, and Tversky (1971) restricted to the preference order over a triangular set of p-gambles (Miyamoto, 1988). Without attempting to state the axioms themselves, we note that a theory implies these axioms if it implies that Eq. (2) or (3) is satisfied, and that the utility function maps the set of outcomes onto an interval of real numbers. The latter is a technical assumption that is satisfied to a close approximation if the domain of the utility function includes a quantity like money whose discrete levels (1¢, 2¢, 3¢, etc.) are virtually continuous increments in subjective value.

The advantage of the generic utility framework is not merely that it is a weak theory, for there are many conceivable weak theories of utility, but rather it is a weak theory whose assumptions are sufficiently strong to provide a basis for the formalization of utility models. Miyamoto (1988) has argued that the principal formalizations in multiattribute utility theory depend on the assumption that the utility measure is an interval scale. In particular, the formalizations of additive, multiplicative, and multilinear utility models by means of utility independence assumptions, and the formalizations of linear/exponential and log/power utility models investigated here all make essential use of the assumption that the measure of utility is an interval scale. A truly generic framework for utility theory must be weak enough to be implied by many other utility theories, but it must also be strong enough to imply an interval scale measure of utility. The important formal result in Miyamoto (1988) was the proof that an interval scale utility measure could be derived from preference assumptions defined only on a triangular set of p-gambles. Previous formalizations required that preference assumptions apply to more general classes of gambles.

Miyamoto and Eraker (1988) presented an experimental study of a multiplicative utility model in the generic utility framework. The stimulus set in this experiment consisted of an upper triangular set of even-chance gambles for multiattribute outcomes. The experiment tested whether preferences for such gambles satisfied a utility independence property that is predicted by additive and multiplicative utility models. Without going into the details of the study, we note that these investigations demonstrate the practical feasibility of investigating standard formalizations of multiattribute utility theory, as developed, for example, in Keeney and Raiffa (1976), within the generic utility framework. Because investigations within the generic utility framework can be interpreted from the standpoint of theories that imply it, these studies demonstrate how methods of multiattribute utility modeling can be introduced into nonstandard utility theories like prospect theory and the dual bilinear model.

Transformational Invariants of Parametric Utility Models

The log/power and linear/exponential utility models are characterized by *transformational invariants* in the sense that each class of models predicts that a property of the response is invariant under a given stimulus transformation. The log/power models predict the invariance $PM(sx\ .5\ sy) = PM(x\ .5\ y)$; the linear/exponential models predict the invariance $PM(x + t\ .5\ y + t) = PM(x\ .5\ y)$. The fact that transformational invariants characterize the linear/exponential and log/power models has been known for a long time (Luce, 1959; Pfanzagl, 1959; Pratt, 1964), but empirical tests of such invariants have been relatively rare in utility analysis. Tversky's (1967) study of the power utility hypothesis for money is an important exception to this claim. He noted that if the utility of a monetary reward is a power function of money, then the certainty match (M) of a gamble must satisfy

$$\log M(x\ p\ 0) = \log x \qquad (1/\omega)\log s(p) \tag{23}$$

where 0 is the outcome where one neither receives nor pays any money, ω is the power parameter for a monetary reward, and $s(p)$ is the subjective probability of the stated probability p. Equation (23) predicts that logarithms of certainty matches are additive functions of x and p in the sense of a two-factor analysis of variance. Tversky (1967) tested Eq. (23) separately in the domain of monetary gains and the domain of monetary losses, and found an excellent fit to the power utility model. In addition to Tversky (1967), a number of studies have investigated the effects of transformations on preference or risk perception (Coombs & Huang, 1970; Keller, Sarin, & Weber, 1986; Payne, Laughhunn, & Crum, 1980, 1981). These studies, however, are not directed at the issue of whether a utility function can be described by a specified parametric model. Outside of preference and utility measurement, there are many studies of the effects of stimulus transformations on perceptual judgments (e.g., Birnbaum & Elmasian, 1977; Shepard & Cooper, 1982).

There are several reasons why experimental tests of transformational invariants are a fruitful approach to utility analysis. First, assuming the validity of the generic utility representation, the empirical validity of the transformational invariants is necessary and sufficient for validity of the corresponding parametric models. Second, tests of transformational invariants are relatively straightforward, and can be performed on the data from individual subjects as well as group data. Third, tests of transformational invariants are informative even when the invariants are violated. The pairwise comparisons of conditions presented in Fig. 2 provide a kind of residual analysis for the parametric utility hypotheses. One can examine the qualitative pattern of violations as well as the frequency or magnitude of violations. Finally, the study of transformational invariants provides a scale-free method for studying the shapes of utility functions. The method is scale-free in the sense that the shape of the utility function can be analyzed without first constructing a numerical utility scale. A scale-free approach has the advantage that information about the shape of the utility function is not confounded with systematic errors in the method of constructing a numerical utility scale (Anderson, 1981; Krantz *et al.*, 1971).

The Utility of Survival Duration

The results presented here demonstrate that the linear/exponential and log/power utility models are systematically violated by the certainty matching judgments of undergraduate subjects. This conclusion must be reached within EU, SEU, ASEU, and NASEU theories, and within the dual bilinear utility model, for these theories all imply that the certainty matches of a triangular set of even-chance gambles satisfy a generic utility representation. Furthermore if the subjects' judgments of their own reference levels of survival duration are valid, the conclusions also follow within prospect theory. With regard to the validity of the judged reference levels, it should be pointed out that the average age of subjects was 20.3 years, and the oldest subject was 28 years old. Even if subjects' judgments of reference levels were invalid, it is plausible that many subjects had reference levels beyond 60 years of age. Such subjects would generally be ERL subjects, because the outcomes in the stimulus gambles ranged from 0 to 36 years. Given the great prevalence of violations of the linear/exponential and log/power models, it is plausible that many violations occurred among subjects whose reference levels were truly external to the range of the gamble outcomes, even if one doubts that the method used in this study actually identified the reference levels of subjects.

The clear rejection of the log/power model found here is in sharp contrast to the large number of studies that have found that subjective representations of quantitative dimensions satisfy a power law (Galanter, 1962; Galanter & Pilner, 1974; Marks, 1974; Stevens, 1957, 1959; Tversky, 1967). There are at least three possible explanations for this divergence. First, the dimension of survival duration may simply be different from other quantitative dimensions. Especially in the near future, each year can be imagined to contain qualitatively different events, e.g., graduate from college, get married, find a job, etc. Perhaps no simple mathematical function represents the effect of these events on the value of survival. Second, it may be that judgments of the worth of gambles are systematically different from the subjective estimates of sensation magnitude on which most studies of the power law have been based. It is certainly the case that the task of judging the certainty matches of gambles is quite different from the tasks of magnitude estimation, ratio estimation, or cross-modality matching (cf. Marks, 1974). Third, it may be that testing transformational invariants is a more powerful technique for detecting departures from the power law than the graphical techniques and goodness of fit measures that have generally been used in studies of the power law. Of course, these explanations are neither exhaustive nor exclusive of each other.

The results of the present experiment cast doubt on the empirical validity of the log/power and linear/exponential models, but we feel that this conclusion should be tempered with several qualifications. First, the applied analysis of medical decisions does not require an exactly valid model of the utility of survival duration. Suppose that there were a utility model that was thought to be exactly valid, and suppose that a second model, which deviated slightly from the first, had parameters that were easier and less expensive to estimate than the parameters of the exact model. If it were found empirically that the optimal therapeutic choice did not generally change when the exact model was replaced by the latter, inexact model, the inexact model would be preferred for practical decision making. It is possible that the linear/exponential models may provide a useful class of inexact utility models for practical decision making. The linear/exponential models were an excellent fit to matching judgments of gambles with middle to long-term survival outcomes. Furthermore, they provided an excellent representation of the results averaged across subjects at all durations of survival. Because the exponential parameter is a simple representation of risk posture, it could prove useful as a representation of risk posture in groups of individuals, for middle to long-range survival outcomes, or as an inexact representation over the entire range of survival.

A second qualification to our conclusions must take note of the fact that they are based entirely on data from certainty matching judgments. Many psychological studies have shown that different methods of measuring preferences yield mutually inconsistent results (Goldstein & Einhorn 1987; Grether & Plott, 1979; Hershey, Kunreuther, & Schoemaker, 1982; Hershey & Schoemaker, 1983; Slovic & Lichtenstein, 1983). We cannot review the evidence concerning preference inconsistencies here, but we will briefly compare certainty matching to preferential choice to show how we approach the problem of inconsistency across methodologies. First, we must define two preference orderings of gambles, denoted $\geqslant_{pc}$ and $\geqslant_{cm}$, that are determined by the methods of preferential choice and certainty matching, respectively. Let p and q be any two probabilities which may or may not be different. The preference ordering determined by preferential choice is found by presenting an individual with pairs of gambles, $(w\ p\ x)$ and $(y\ q\ z)$, and asking the individual to decide which is the more desirable gamble. Let $(w\ p\ x) \geqslant_{pc} (y\ q\ z)$ indicate that the individual regards $(w\ p\ x)$ to be equal or more desirable than $(y\ q\ z)$ in a preferential choice task. The preference ordering determined by certainty matching is found by asking an individual to judge the certainty matches of gambles, and then ordering the gambles by the magnitude of the matches. If $M(w\ p\ x)$ and $M(y\ q\ z)$ are the certainty matches of $(w\ p\ x)$ and $(y\ q\ z)$, respectively, let $(w\ p\ x) \geqslant_{cm} (y\ q\ z)$ indicate that $M(w\ p\ x) \geqslant M(y\ q\ z)$. Although logical consistency would require that one always give a higher certainty match to the gamble which one would choose in a preferential choice, experimental evidence clearly shows that individuals do not generally behave this way (Goldstein & Einhorn, 1987; Grether & Plott, 1979; Lichtenstein & Slovic, 1971; Lindman, 1971; Slovic & Lichtenstein, 1983). In other words, it has been found empirically that $\geqslant_{pc}$ and $\geqslant_{cm}$ are not identical orderings of gambles. This empirical finding has sometimes been referred to as the *preference reversal phenomenon* because some of the preferences determined by one method are reversed when a different method is used.

Because of the existence of preference reversals, one cannot assume that utility analyses carried out with one method of preference measurement, e.g., certainty matching, generalize to results obtained by a second method of preference measurement, e.g., preferential choice. We would also argue, however, that the inconsistencies do not rule out the possibility that utility analyses generalize across methods. To see why this is so, suppose that the two preference orderings, $\geqslant_{pc}$ and $\geqslant_{cm}$, are both represented by subjectively weighted averages:

$(x\ p\ y) \geqslant_{pc}$

iff $s_1(p)U\ \ (x) + \ldots\ p)U\ \ s_1(q)U\ \ldots\ q)U$ (24)

$(x\ p\ y) \geqslant_{cm} (w\ q$

iff $s_2(p)U_2(x)\ \ldots\ p)U\ \ldots_2(q)U_2(w\ \ldots\ q)U$ (25)

In these representations, the s_i scales represent subjective probabilities and the U_i scales represent the utility of outcomes ($i = 1,2$). The subjective probability and utility scales are subscripted in order to show that they may differ in the representation of preferential choice and certainty matching.

Assuming for the sake of discussion that (24) and (25) are valid, the existence of preference reversals implies that either s_1 and s_2 are not linearly related, or U_1 and U_2 are not linearly related, or both. From the standpoint of utility analysis, however, only a nonlinearity between U_1 and U_2 is of concern. Since a utility scale is an interval scale, utility scales that are linearly related are empirically equivalent in the sense that their observable implications are identical. Thus, if U_1 and U_2 are linearly related, precisely the same utility models are valid relative to preferential choice and certainty matching, even if the preference *orderings*, $\geqslant_{pc}$ and $\geqslant_{cm}$, are different. Notice that in the context of medical decision analysis, it is perfectly natural to be interested exclusively in the utility functions of subjects, and not their subjective probabilities. The probabilities of outcomes are typically assessed from expert opinion, that is, the opinions of physicians and other health professionals (Weinstein *et al.*, 1980). The utilities, however, must come from patients or more accessible lay people who are serving as experimental subjects, as in the present study. Insofar as our goal is to model the utilities of such individuals for survival duration or other health attributes, it is of no direct concern if their subjective probability representations are altered by different methods of preference measurement as long as their utility functions are not also affected by method of measurement.

In effect, we are suggesting that inconsistencies across methods of preference measurement may be due solely to changes in the representation of subject probability in the various methods. Existing evidence neither suggests that this is the case, nor rules it out. Until further research determines whether utility functions change nonlinearly under different methods of preference measurement, one cannot assume that utility analyses carried out using one method generalize or fail to generalize to other methods of preference measurement.

Summary

The log/power and linear/exponential utility functio were investi

gated in a framework called generic utility theory, whose assumptions are implied by other important utility theories including EU and SEU theory, prospect theory, and the dual bilinear model. The generic utility theory is a useful working framework for utility analysis because analyses developed within this framework are interpretable from the standpoint of stronger theories that imply its assumptions, without commiting one to the complete set of assumptions underlying stronger theories. Standard expected utility formalizations of the log/power and linear/exponential models were reformulated in the generic utility framework, and then tested experimentally in terms of certainty matching judgments of gambles for survival duration. The log/power and linear/exponential models predict that proportional matches, which are a transformation of certainty matching judgments, are invariant under multiplication by a constant and addition of a constant, respectively. Both predictions were rejected by a substantial majority of subjects. A primary motivation for investigating log/power and linear/exponential utility models is that both classes of models have parameters that constitute simple representations of risk posture. Risk posture with respect to survival duration is known to be critically important in analyses of medical therapeutic decisions when therapeutic alternatives differ in their trade-offs between short- and long-term survival. Although the results show that these parametric classes of models are not valid as exact descriptions of individual utility functions, they may still be useful as approximate representations in applied decision making. In particular, the linear/exponential utility models were an excellent approximation for results averaged across individuals, and for middle to long-term survival outcomes.

APPENDIX

The equivalence between Eqs. (15) and (18) is proved as follows: for any x, y and s that is not too large,

$$M(sx\ .5\ sy)\quad sM(x\ .5\ y) \qquad (15$$

iff

$$M(sx\ .5\ s\ldots\quad sx\quad sM(x\ .5\ y)\quad sx$$

iff

$$\frac{M(sx\ .5\ sy) - sx}{sy - sx}\quad \frac{sM(x\ .5\ y) - sx}{sy - sx}$$

iff

$$PM(sx\ .5\ sy)\quad PM(x\ .5\ \ldots \qquad (18$$

Therefore (15) is equivalent to (18).

The equivalence between Eqs. (16) and (19) is proved as follows: for any x, y, and t that is not too large,

$$M(x + t\ .5\ y + t) = M(x\ .5\ y) + t \tag{16}$$
iff
$$M(x + t\ .5\ y + t) - (x + t) = M(x\ .5\ y) + t - (x + t)$$
iff
$$\frac{M(x + t\ .5\ y + t) - (x + t)}{(y + t) - (x + t)} = \frac{M(x\ .5\ y) - x}{y - x}$$
iff
$$PM(x + t\ .5\ y + t) = PM(x\ .5\ y). \tag{19}$$

Therefore (16) is satisfied iff (19) is satisfied.

REFERENCES

Aczel, J. (1966). *Lectures on functional equations and their applications*. New York: Academic Press.

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

Birnbaum, M. H., & Elmasian, R. (1977). Loudness "ratios" and "differences" involve the same psychophysical operation. *Perception & Psychophysics,* **22,** 383–391.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice–Hall.

Coombs, C. H., & Huang, L. C. (1970). Polynomial psychophysics of risk. *Journal of Mathematical Psychology,* **7,** 317–338.

Coombs, C. H., & Huang, L. C. (1976). Tests of the betweenness property of expected utility. *Journal of Mathematical Psychology,* **13,** 323–337.

Debreu, G. (1959). Cardinal utility of even-chance mixtures of pairs of sure prospects. *Review of Economic Studies,* **71,** 174–177.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review,* **69,** 109–135.

Galanter, E. (1962). The direct measurement of utility and subject probability. *American Journal of Psychology,* **75,** 208–220.

Galanter, E., & Pliner, P. (1974). Cross-modality matching of money against other continua. In H. R. Moskowitz, B. Scharf, & J. C. Stevens (Eds.), *Sensation and measurement (pp. 65–76). Dordrecht: Reidel.*

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review,* **94,** 236–254.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review,* **69,** 623–638.

Hershey, J. C., Kunreuther. H. C., & Schoemaker, P. J. H. (1982). Sources of bias in assessment procedures for utility functions. *Management Science* **28,** 936–954.

Hershey, J. C., & Schoemaker, P. J. H. (1983). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science,* **31,** 1213–1231.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica,* **47,** 263–291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist,* **39,** 341–350.

Karmarkar, U. S. (1978). Subjectively weighted utility: A descriptive extension of the expected utility model. *Organizational Behavior and Human Performance,* **21,** 61–72.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Keller, L. R., Sarin, R. K., & Weber, M. (1986). Empirical investigation of some properties of the perceived riskiness of gambles. *Organizational Behavior and Human Decision Processes,* **38,** 114–130.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.

Lichtenstein, S., & Slovic, P. (1971). Reversal of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology,* **89,** 46–55.

Lindman, H. R. (1971). Inconsistent preferences among gambles. *Journal of Experimental Psychology,* **89,** 390–397.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review,* **66,** 81–95.

Luce, R. D., & Narens, L. (1985). Classification of concatenation measurement structures according to scale type. *Journal of Mathematical Psychology,* **29,** 1–72.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York: Wiley.

Marks, L. E. (1974). *Sensory processes: The new psychophysics*. New York: Academic Press.

McClelland, G. H., & Coombs, C. H. (1975). ORDMET: A general algorithm for constructing all numerical solutions to ordered metric structures. *Psychometrika,* **40,** 269–290.

McNeil, B. J., & Pauker, S. G. (1982). Optimizing patient and societal decision making by the incorporation of individual values. In R. L. Kane & R. A. Kane (Eds.), *Values and long-term care* (pp. 215–230). Lexington, MA: D.C. Health.

McNeil, B. J., Weichselbaum, R., & Pauker, S. G. (1978). Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine,* **299,** 1397–1401.

Miyamoto, J. M. (1985). *The utility of survival duration and health quality: A conjoint measurement analysis* (Publication No 8512473). Ann Arbor, MI: University Microfilms, Inc.

Miyamoto, J. M. (1987a). Constraints on the representation of gambles in prospect theory. *Journal of Mathematical Psychology,* **31,** 410–418.

Miyamoto, J. M. (1987b). *The effect of health quality on the utility of survival duration.* Paper presented at a Symposium on Psychological and Economic Perspectives on Choices between Present and Future Outcomes. Annual Meeting of the Society for Judgment and Decision Making, Seattle, WA.

Miyamoto, J. M. (1988). Generic utility theory: Measurement foundations and applications in multiattribute utility theory. *Journal of Mathematical Psychology,* **32,** 357–404.

Miyamoto, J. M., & Eraker, S. A. (1985). Parameter estimates for a QALY utility model. *Medical Decision Making,* **5,** 191–213.

Miyamoto, J. M., & Eraker, S. A. (1988). A multiplicative model of the utility of survival duration and health quality. *Journal of Experimental Psychology: General,* **117,** 3–20.

Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin,* **99,** 166–180.

Pauker, S. G., & McNeil, B. J. (1981). Impact of patient preferences on the selection of therapy. *Journal of Chronic Disease,* **34,** 77–86.

Payne, J. W., Laughhunn, D. J., & Crum, R. (1980). Translation of gambles and aspiration level effects in risky choice behavior. *Management Sciences,* **26,** 1039–1060.

Payne, J. W., Laughhunn, D. J., & Crum, R. (1981). Further tests of aspiration level effects in risky choice behavior. *Management Science,* **27,** 953–958.

Pfanzagl, J. (1959). A general theory of measurement: Applications to utility. *Naval Research Logistic Quarterly,* **6,** 283–294.

Pliskin, J. S., Stason, W. B., Weinstein, M. C., Johnson, R. A., Cohn, P. F., McEnany

M. T., & Braun, P. (1981). Coronary artery bypass graft surgery: Clinical decision making and cost-effectiveness analysis. *Medical Decision Making*, **1**, 10–28.

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research*, **28**, 206–224.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, **32**, 122–136.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press/Bradford Books.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review*, **73**, 569–605.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153–181.

Stevens, S. S. (1959). Measurement, psychophysics, and utility. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 18–63). New York: Wiley.

Tversky, A. (1967). Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology*, **75**, 27–36.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the rationality of choice. *Science*, **211**, 453–458.

Von Winterfeldt, D. V., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, UK: Cambridge, Univ. Press.

Weinstein, M. C., Fineberg, H. V. Elstein, A. S., Frazier, H. S., Neuhauser, D., Neutra, R. R., & McNeil, B. J. (1980). *Clinical decision analysis*. Philadelphia: Saunders.

Weinstein, M. C., Pliskin, J. S., & Stason, W. B. (1977). Coronary artery bypass surgery: Decision and policy analysis. In J. P. Bunker, B. A. Barnes, & F. Mosteller (Eds.), *Costs, risks, and benefits of surgery* (pp. 342–371). New York: Oxford Univ. Press.

Weinstein, M. C., & Stason, W. B. (1982). Cost-effectiveness of coronary artery bypass surgery. *Circulation*, **66:III**, 56–66.

RECEIVED: April 29, 1987