Cognitive Psychology **38,** 129–166 (1999)
Article ID cogp.1998.0710, available online at http://www.idealibrary.com on IDEAL®

On the Shape of the Probability Weighting Function

Richard Gonzalez

University of Michigan

and

George Wu

University of Chicago, Graduate School of Business

Empirical studies have shown that decision makers do not usually treat probabilities linearly. Instead, people tend to overweight small probabilities and underweight large probabilities. One way to model such distortions in decision making under risk is through a probability weighting function. We present a nonparametric estimation procedure for assessing the probability weighting function and value function at the level of the individual subject. The evidence in the domain of gains supports a two-parameter weighting function, where each parameter is given a psychological interpretation: one parameter measures how the decision maker discriminates probabilities, and the other parameter measures how attractive the decision maker views gambling. These findings are consistent with a growing body of empirical and theoretical work attempting to establish a psychological rationale for the probability weighting function. © 1999 Academic Press

The perception of probability has a psychophysics all its own. If men have a 2% chance of contracting a particular disease and women have a 1% chance, we perceive the risk for men as twice the risk for women. However, the same difference of 1% appears less dramatic when the chance of con-

This research was supported by a grant from the National Science Foundation (R.G.) and by the James S. Kemper Foundation Faculty Research Fund at the Graduate School of Business, the University of Chicago (G.W.). Portions of these data were presented at the Society for Mathematical Psychology Meetings, Norman, OK, 1993. We thank John Miyamoto and Peter Wakker for their insightful discussions about weighting functions. Amos Tversky provided helpful guidance during the early stages of this research. Lyle Brenner, Dale Griffin, Chip Heath, Josh Klayman, Duncan Luce, Drazen Prelec, Eldar Shafir, and Peter Wakker provided helpful comments on a previous draft.

Address correspondence and reprint requests to either author: Richard Gonzalez, Department of Psychology, University of Michigan, Ann Arbor, MI 48109, or George Wu, University of Chicago, Graduate School of Business, 1101 E. 58th Street, Chicago, IL 60637. E-mail: either gonzo@umich.edu or george.wu@gsb.uchicago.edu.

0010-0285/99 $30.00

tracting the disease is near the middle of the probability scale, e.g., a 33% chance for men and a 32% chance for women may be perceived as a trivial sex difference.

Consider a second example. Suppose a researcher is deliberating whether to spend a day performing additional analyses for a manuscript. The researcher believes that these analyses will improve the probability of acceptance by 5%. We suggest that the author is more likely to perform the additional analyses if she believes the manuscript has a 90% chance of acceptance than if she regards her chances as 30%. Put differently, improving her chances from 90 to 95% seems more substantial than improving from 30 to 35%.

In both of these examples the impact of additional probability depends on whether it is added to a small, medium, or large probability (for related examples see Quattrone & Tversky, 1988). We collected survey data as a preliminary test of these intuitions.[1] Fifty-six undergraduates were given the following question:

> You have two lotteries to win $250. One offers a 5% chance to win the prize and the other offers a 30% chance to win the prize.
> A: You can improve the chances of winning the first lottery from 5 to 10%.
> B: You can improve the chances of winning the second lottery from 30 to 35%.
> Which of these two improvements, or increases, seems like a more significant change? (circle one)

The majority of respondents (75%) viewed option A as the more significant improvement. Even though the dependent variable is not a standard choice task, these data can be interpreted as respondents' self-report that a change from 5 to 10% is seen as a more significant increase than a change from 30 to 35%.

The same respondents were also given a different question where the stimulus probabilities were translated by .60:

> You have two lotteries to win $250. One offers a 65% chance to win the prize and the other offers a 90% chance to win the prize.
> C: You can improve the chances of winning the first lottery from 65 to 70%.
> D: You can improve the chances of winning the second lottery from 90 to 95%.
> Which of these two improvements, or increases, seems like a more significant change? (circle one)

In the second question, only 37% of the participants viewed option C as a more significant improvement. The modal choice of A and D suggests that a change from .05 to .10 is seen as more dramatic than a change from .30 to .35, but a change from .65 to .70 is viewed as less significant than a change from .90 to .95. The difference between the choice proportions in the two

[1] Data for this survey were collected in collaboration with Amos Tversky.

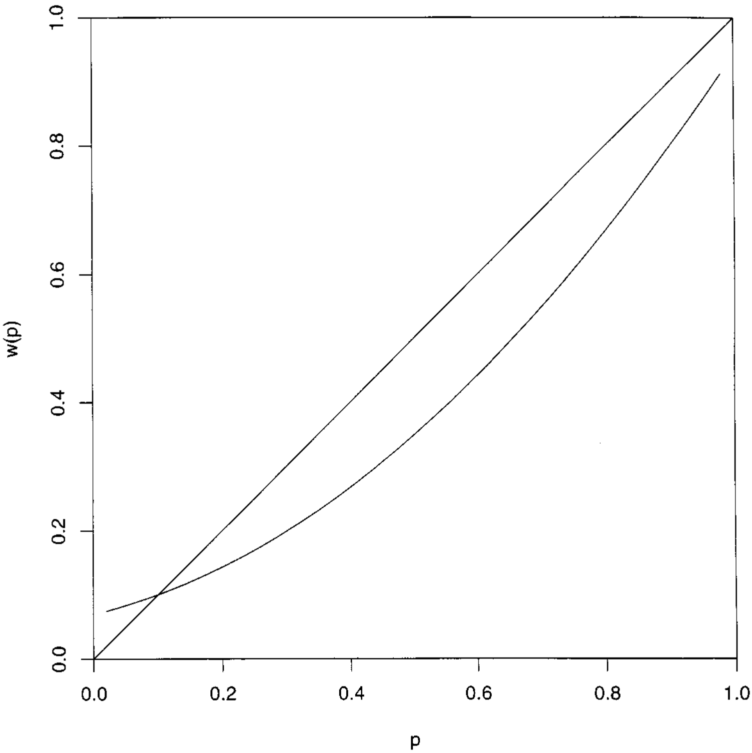

FIG. 1. Weighting function proposed in Prospect Theory (Kahneman & Tversky, 1979), which is not defined near the end points. The key properties are the overweighting of small probability and the underweighting of large probability.

problems is statistically significant by McNemar's test, $\chi^2(1) = 19.2$, $p <$ .0001. The order of the two questions was counterbalanced.

Taken together, the two examples and the survey questions suggest that individuals do not treat probabilities linearly. In this paper we present evidence based on a more traditional choice task that is consistent with this informal observation. This idea is modeled formally in prospect theory, which permits a probability distortion through a *probability weighting function.* Kahneman and Tversky (1979) presented a stylized probability weighting function (see Fig. 1) that exhibited a set of basic properties meant to organize empirical departures from classical expected utility theory. Perhaps the two most notable properties of Figure 1 are the overweighting of small probabilities and the underweighting of large probabilities. We denote the probability weighting function by $w(p)$, a function that maps the [0,1] interval onto itself. It is important to note that the weighting function is not a subjective probability but rather a distortion of the given probability (see

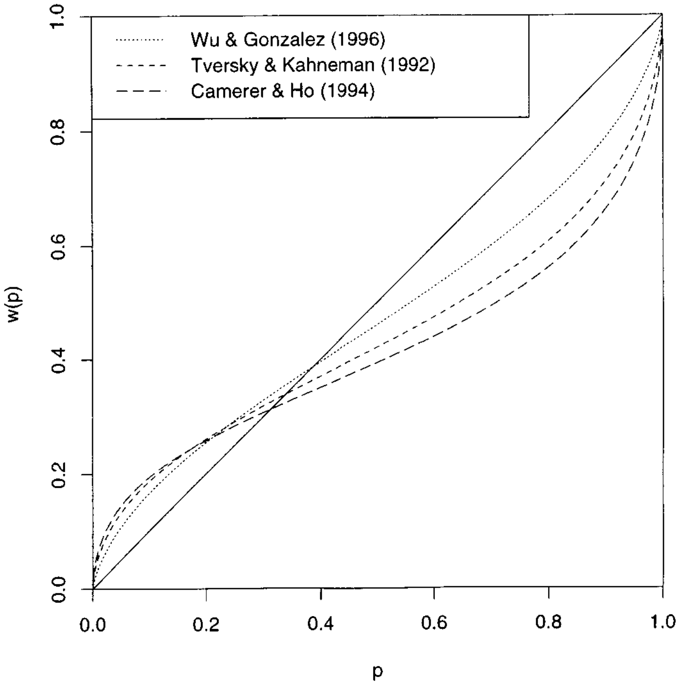

FIG. 2. One-parameter weighting functions estimated by Camerer and Ho (1994), Tversky and Kahneman (1992), and Wu and Gonzalez (1996) using $w(p) = (p^{\beta}/(p^{\beta} + (1 - p)^{\beta})^{1/\beta})$. The parameter estimates were .56, .61, and .71, respectively.

Kahneman & Tversky, 1979). An individual may agree that the probability of a fair coin landing on heads is .5, but in decision making distort that probability by $w(.5)$.

The weighting function shown in Fig. 1 cannot account for the pattern discussed above because it is not concave for low probability. The introductory examples suggest that probability changes appear more dramatic near the endpoints 0 and 1 than near the middle of the probability scale. Generalized, this implies a probability weighting function that is inverse-S-shaped: concave for low probability and convex for high probability. Weighting functions consistent with the survey data are shown in Fig. 2; empirical support for this shape appeared in three recent choice studies (Camerer & Ho, 1994; Hartinger, 1998; Tversky & Kahneman, 1992; Wu & Gonzalez, 1996).

The general question studied in this paper is how the psychophysics of probability influences decision making under risk. In turn, an understanding of the probability weighting function will provide insights about the psychology of risk. The outline of the paper is as follows: we first provide a sketch of the relevant theoretical background, review relevant studies, and discuss the limitations of those studies. We then review a psychological rationale for the shape of the weighting function in terms of discriminability and attractiveness and suggest a functional form that can model these psychological intuitions. Next, we present a new study and a nonparametric algorithm that permits the estimation of individual subjects' value and weighting functions in a manner that eliminates many of the shortcomings of previous work. Data at both the aggregate level and the individual level are consistent with the inverse-S-shape weighting function. We conclude by discussing the implication of these results for future research and for applied settings.

Modeling Probability Distortions: Prospect Theory

Preston and Baratta (1948) made an early contribution toward modeling probability distortions. They used gambles with one nonzero outcome such as (100, .25; 0, .75). This notation represents the gamble offering a 25% chance to win \$100 and a 75% chance to win \$0. They collected certainty equivalents (actually, buying prices) in the context of an incentive compatible procedure. A certainty equivalent, denoted CE, is the amount of money for which a person is indifferent between receiving that amount of money for certain or playing the gamble. Preston and Baratta assumed a separable representation with a linear value function, $v(X) = X$, in the sense that CE $= w(p)X$ for the gamble $(X, p; 0, 1 - p)$. They observed that the weighting function w (estimated under linear v) was *regressive;* that is, there appeared to be overweighting relative to the identity line for $p = .01$ and $p = .05$ and underweighting for p's in the set {.25, .50, .75, .95, .99}.

There are two problems with their analyses. First, the analysis assumes a linear value function. Even though ''duals'' to expected utility with linear v have been proposed (e.g., Yaari, 1987; Weibull, 1982) and one of the early resolutions to the St. Petersburg paradox used nonlinear w and linear v (see Arrow, 1951, for a discussion), a nonlinear utility function typically provides a better fit for both risk and nonrisk domains (e.g., Fishburn & Kochenberger, 1979; Galanter & Pliner, 1974; Parker & Schneider, 1988; Tversky & Kahneman, 1992). Second, even if Preston and Baratta had used a nonlinear v they would have been unable to extract a unique estimate of the weighting function because their one nonzero outcome stimuli did not permit separation of the weighting function from the value function. Estimates of v and w using one nonzero outcome stimuli are unique only to a power (i.e., if v and w represent preferences then so will v^{α} and w^{α}). Gambles with at least two nonzero outcomes are required to separate v and w. Methodological problems aside, Preston and Baratta's results captured the general flavor of the weighting function: regressive with a crossover point near $p = .30$. Similar qualitative results of nonlinear w were obtained by Mosteller and Nogee (1951) and Edwards (1954).

Kahneman and Tversky (1979) took a different approach. They identified properties of w that would accommodate a set of general departures from expected utility behavior, specifically, the common ratio effect (choices change when the probabilities in a pair of gambles are scaled by a common factor) and the common consequence effect (choices change when probability mass in a pair of gambles is shifted from one common consequence to another). [For details on how generalizations of these effects relate to the probability weighting function see Prelec (1998) and Wu & Gonzalez (1998), respectively (see also Tversky & Wakker, 1995).] The properties of the weighting function identified by Kahneman and Tversky included overweighting of small probabilities, underweighting of large probabilities, and *subcertainty* (i.e., the sum of the weights for complementary probabilities is less than one, $w(p) + w(1 - p) < 1$). Kahneman and Tversky also noted that the probability weighting function may not be well behaved near the endpoints 0 and 1. The function shown in Fig. 1 is consistent with these properties.

Kahneman and Tversky (1979) recognized that introducing a nonlinear weighting function without further modification of the model would lead to predicted violations of stochastic dominance.[2] Violations of stochastic dominance are difficult to observe empirically unless the stochastic dominance is not transparent (Tversky & Kahneman, 1986; Birnbaum, 1997; Leland, 1998).

To avoid predictions of transparent stochastic dominance violations, Kahneman and Tversky proposed a number of editing operations. Here we consider only the case of two outcome gambles involving either all gains or all losses. Consider a gamble offering a 50% chance to win \$100 and a 50% chance to win \$25. The psychological intuition for one of their editing rule is as follows: regardless of how the chance event plays out, the gamble is sure to offer at least \$25, plus a 50% chance of receiving an additional amount. This intuition can be represented symbolically

$$v(Y) + w(p)[v(X) - v(Y)], \tag{1}$$

where v is the value function, w is the weighting function, and for this example $X = 100$, $Y = 25$, and $p = .50$.

More recently, Tversky and Kahneman (1992) generalized prospect theory using a rank-dependent, or cumulative, representation (see also Quiggin, 1993; Luce & Fishburn, 1991, 1995; Starmer & Sugden, 1989; Wakker & Tversky, 1993). Intuitively, cumulative prospect theory (CPT) generalizes the idea of the editing rule exhibited in Eq. (1) to gambles with an arbitrary number of outcomes. Thus, cumulative prospect theory does not require ex-

[2] Denote the cumulative probability that gamble I offers for a given outcome X as $F_I(X)$. Gamble A stochastically dominates gamble B iff $F_A(X) \leq F_B(X)$ for all X and $F_A(X) < F_B(X)$ for at least one X. For a more extended discussion of the problems surrounding violations of stochastic dominance see Fishburn (1978) and Quiggin (1993).

plicit editing operations in order to avoid predicted violations of stochastic dominance (however, see Wu, 1994).

Much of the technical generalization of prospect theory involves the combination rule of how the value function is combined with the probability weighting function. CPT consists of the sum of two rank-dependent expected utility representations, one for gains and one for losses. In the special case in which all outcomes are nonnegative, the representation of the gamble $(X, p; Y, 1 - p)$ under CPT where $X > Y \geq 0$ is

$$w(p)v(X) + [1 - w(p)]v(Y), \tag{2}$$

where w is a probability weighting function and v is a value function. The term ‘‘rank dependent’’ applies because the weight attached to an outcome ($w(p)$ for the higher outcome and $1 - w(p)$ for the lower outcome) depends on the rank of that outcome with respect to other outcomes in the gamble (see Quiggin, 1982; Yaari, 1987; Segal, 1989; Wakker, 1989, 1994). Note that Eq. (2) is, after algebraic rearrangement, identical to Eq. (1). One advantage of the rank-dependent approach is that Eq. (2) is easily generalized to n-outcome gambles. For gambles of the form $(X_1, p_1; \ldots; X_n, p_n)$, where $|X_i| > |X_{i+1}|$ and all X's are on the same side of the reference point, the representation is

$$w(p_1)v(X_1) + \sum_{i=2}^{n} \left[w\left(\sum_{j=1}^{i} p_j \right) - w\left(\sum_{j=1}^{i-1} p_j \right) \right] v(X_i).$$

In the context of cumulative prospect theory (and rank-dependent utility models) it is useful to distinguish the probability weighting function from the ‘‘decision weight.’’ The probability weighting function models the distortion of probability (i.e., $w(p)$) and characterizes the psychophysics of chance. The decision weight is the term that multiplies the value of each outcome. Note that the rank-dependent intuition applies here in the sense that the value of the highest outcome is weighted by $w(p)$, and all other values for $i > 1$ are weighted by decision weights of the form

$$w\left(p_i + \sum_{j=1}^{i-1} p_j \right) - w\left(\sum_{j=1}^{i-1} p_j \right).$$

Thus, in a two-outcome gamble with $p_1 = p_2 = .5$, the decision weight attached to each of the two outcomes will differ. As a further illustration, consider a gamble offering 50% chance to win \$100, \$50 otherwise. According to the rank-dependent model, an individual with $w(.5) = .3$ will weight $v(100)$ by .3 and $v(50)$ by .7. However, the same individual offered a 50% chance to win \$200 and \$100 otherwise, will weight $v(200)$ by .3

and $v(100)$ by .7. Thus, the probability weighting function captures the psychology of probability distortion, whereas how the outcomes are weighted depends on the particular combination rule, such as CPT.

In sum, the history of research attempting to describe how people make decisions in domains of risk can be characterized by the following questions: Do people distort outcomes and how? Do people distort probabilities and how? How should these distortions be interpreted and how do they inform us about how people choose among risky alternatives? How can we distinguish distortions of probability from distortions of outcomes? How do we combine outcomes with probabilities to model decision? CPT is an attempt to address most of these questions. A question we have not yet addressed is the psychological interpretation of the weighting function.

Psychological Interpretation of the Weighting Function

In this section we discuss two features of the weighting function that can be given a psychological interpretation. One feature involves the degree of curvature of the weighting function, which can be interpreted as *discriminability*, and the other feature involves the elevation of the weighting function, which can be interpreted as *attractiveness*. Similar but weaker concepts (source sensitivity and source preference) in the context of decision making under uncertainty were proposed by Tversky and Wakker (1995) and empirically tested by Tversky and Fox (1995).

Diminishing sensitivity and discriminability. There was relatively little progress in establishing a psychological foundation for the weighting function until Tversky and Kahneman (1992) offered a psychological hypothesis. The notion, which Tversky and Kahneman called *diminishing sensitivity*, was very simple: people become less sensitive to changes in probability as they move away from a reference point. In the probability domain, the two endpoints 0 and 1 serve as reference points in the sense that one end represents ‘‘certainly will not happen’’ and the other end represents ‘‘certainly will happen.’’ Under the principle of diminishing sensitivity, increments near the end points of the probability scale loom larger than increments near the middle of the scale. Diminishing sensitivity also applies in the domain of outcomes with the status quo usually serving as a single reference point.

Diminishing sensitivity suggests that the weighting function has an inverse-S-shape—first concave and then convex. That is, sensitivity to changes in probability decreases as probability moves away from the reference point of 0 or away from the reference point of 1. This inverse-S-shaped weighting function can account for the results of Preston and Baratta (1948) and Kahneman and Tversky (1979). Evidence for an inverse-S-shaped weighting function was also found in aggregate data by Camerer and Ho (1994) using very limited stimuli designed to test betweenness, by Tversky and

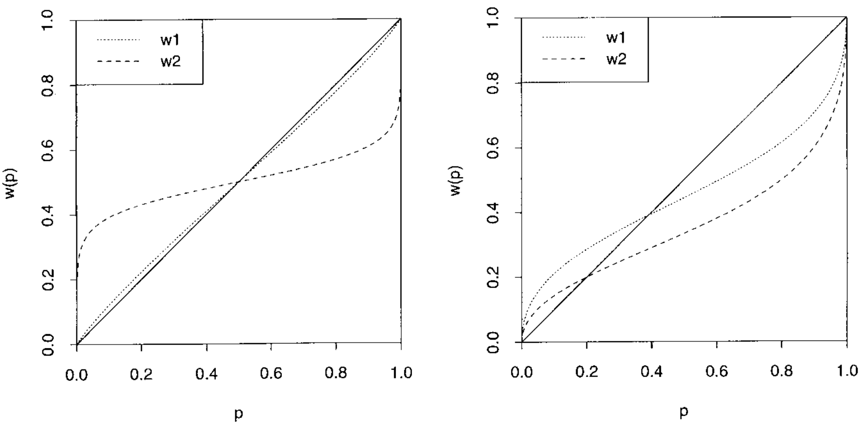

FIG. 3. (Left) Two weighting functions that differ primarily in curvature—w_1 is relatively linear and w_2 is almost a step function. (Right) Two weighting functions that differ primarily in elevation—w_1 overweights relative to w_2.

Kahneman (1992) using certainty equivalence data, by Wu and Gonzalez (1996) using series of gambles constructed to examine the curvature of the weighting function, and by Abdellaoui (1998) using a nonparametric estimation task. A weaker condition, bounded subadditivity, was also supported by Tversky and Fox (1995). For a counterexample see Birnbaum and McIntosh (1996).

Diminishing sensitivity is related to the concept of discriminability in the psychophysics literature in the sense that the sensitivity to a unit difference in probability changes along the probability scale. *Discriminability* may be characterized as follows: weighting function w_1 is said to exhibit greater discriminability, or sensitivity, than weighting function w_2 within interval $[q_1, q_2]$ whenever $w_1(p + \epsilon) - w_1(p) > w_2(p + \epsilon) - w_2(p)$ for all p bounded away from 0 and 1, $\epsilon > 0$, and $p, p + \epsilon \in [q_1, q_2]$. That is, changes (first-order differences) within an interval along w_1 are more pronounced than changes along w_2. The boundary conditions are needed because $w(0) = 0$ and $w(1) = 1$ by definition, and for any continuous weighting function the following property holds: $\int_0^1 w'(p)dp = 1$.

Discriminability can be illustrated by considering two extreme cases: a function that approaches a step function and a function that is almost linear (see the left panel of Fig. 3). The step function shows less sensitivity to changes in probability than the linear function, except near 0 and 1. A step function corresponds to the case in which an individual detects ‘‘certainly will’’ and ‘‘certainly will not,’’ but all other probability levels are treated equally (such as the generic ‘‘maybe’’). Piaget and Inhelder (1975) observed that a 4-year-old child’s understanding of chance corresponds to this type of step function. In contrast, a linear weighting function exhibits more (and constant) sensitivity to changes in probability than a step function. Two studies suggest that some experts possess relatively linear weighting functions when gambling in their domain of expertise: the efficiency of parimutuel betting markets suggests that many racetrack betters are sensitive to small differences in odds (see, e.g., Thaler & Ziemba, 1988), and a study of options traders found that the median options trader is an expected value maximizer and thus shows equal sensitivity throughout the probability interval (Fox, Rogers, & Tversky, 1996). Discriminability could also be defined intrapersonally, e.g., an option trader may exhibit more discriminability for gambles based on options trading rather than gambles based on horse races.

Attractiveness. While useful, the concept of diminishing sensitivity provides an incomplete account of the weighting function. Even though the concept can explain the curvature of the weighting function, it cannot account for the level of absolute weights. That is, diminishing sensitivity merely predicts that w is first concave and then convex. But the property is silent about underweighting or overweighting relative to the objective probability (i.e., the 45 degree line). An inverse-S-shaped weighting function can be completely below the identity line, can cross the identity line at some point, or can be completely above the identity line—everywhere maintaining its concave–convex shape.

Thus, a second feature of the probability weighting function corresponds to the absolute level of w. For example, consider two people who each face a 50% chance to win \$$X$ (\$0 otherwise). One person's weighting function yields $w_1(.5) = .6$ whereas the other yields $w_2(.5) = .4$; then we say that the first person finds the gamble more ''attractive'' because he assigns a greater weight to the probability .5.

This concept can be generalized in terms of the elevation of the weighting function in the w vs. p plot. If for all p, individual 1 assigns more weight to p than individual 2, i.e., $w_1(p) \geq w_2(p)$ for all p with at least one strict inequality, then individual 1's w graph is ''elevated'' relative to individual 2's graph (see the right panel of Fig. 3). Note that elevation is logically independent of curvature. In the context of CPT with two outcome gambles (all gains), a weighting function that is more elevated will assign a greater weight to the higher outcome.

We interpret this interpersonal difference in elevation as a *attractiveness;* i.e., one person finds betting on the chance domain more attractive than the second person. An analogous definition could also be given for intrapersonal comparisons of two different chance domains: a person finds chance domain 1 more attractive than chance domain 2 iff $w_1(p) \geq w_2(p)$ for all p. For example, a person may prefer to bet on sporting events rather than on the outcomes of political elections holding constant the chance winning (Heath & Tversky, 1991), or a person may prefer a lottery in which she is able to select her own numbers to one in which numbers are assigned to her (such as in the illusion of control work by Langer, 1975). Note that we interpret the illusion of control as due to the probability weighting function rather than due to differences in subjective probability or the value function.

In sum, there appear to be two logically independent psychological properties that characterize the weighting function. Discriminability refers to how people discriminate probabilities in an interval bounded away from 0 and 1. Attractiveness refers to the degree of over/under weighting. The former property is indexed by the curvature of the weighting function, and the latter is indexed by the elevation of the weighting function.[3]

Functional Forms for w

If there are two logically independent, psychological properties to the weighting function, then it should be possible to model w with two parameters such that one parameter represents curvature (discriminability) and the other parameter represents elevation (attractiveness). One way to derive such a two parameter w is to note that on the log odds scale a linear transformation can be used to vary elevation (intercept) and curvature (slope) separately, i.e.,

$$\log \frac{w(p)}{1 - w(p)} = \gamma \log \frac{p}{1 - p} + \tau.$$

Solving for $w(p)$ we get

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1 - p)^{\gamma}}, \tag{3}$$

where $\delta = \exp \tau$. In Eq. (3), the γ parameter primarily controls curvature and δ primarily controls elevation. We call the functional form in Eq. (3) ''linear in log odds.'' It is a variant of the form used by Lattimore, Baker, and Witte (1992) and was used by Goldstein and Einhorn (1987), Tversky and Fox (1995), Birnbaum and McIntosh (1996), and Kilka and Weber (1998). Karmarkar (1978, 1979) used the special case of the linear in log odds form with $\delta = 1$. A preference condition that in the context of rank-

[3] Lopes (1987, 1990) used the phrases ''security-minded'' for w that is convex and everywhere below the identity line, ''potential-minded'' for w that is concave and everywhere above the identity line, and ''cautiously-hopeful'' for the inverse-S-shaped w depicted in Fig. 2 (see also Weber, 1994, for a discussion). The present concepts of discriminability and attractiveness provide a more detailed account of the weighting function, without confounding curvature and elevation. The Lopes framework also includes the concept of aspiration level, which we do not model here.

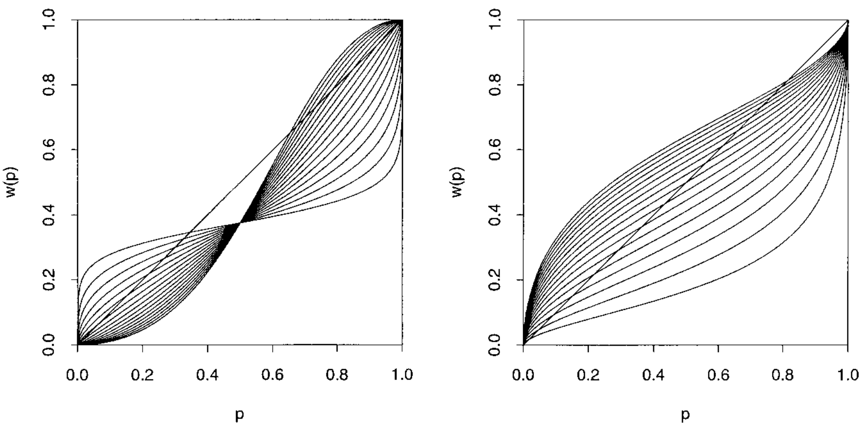

FIG. 4. Demonstration that γ primarily controls curvature and δ primarily controls elevation (parameters from Eq. (3)). The first panel fixes δ at .6 and varies γ between .2 and 1.8. The second panel fixes γ at .6 and varies δ between .2 and 1.8.

dependent theory is necessary and sufficient for the linear in log odds form is given in the Appendix.

Figure 4 shows how the two parameters control curvature and elevation almost independently. The first panel holds δ fixed at .6 and varies γ between .2 and 1.8 in increments of .1. The second panel holds γ fixed at .6 and varies δ between .2 and 1.8. Note that γ essentially controls the degree of curvature, and δ essentially controls the elevation. Because the weighting function is constrained at the end points ($w(0) = 0$ and $w(1) = 1$), an independent separation of curvature and elevation is not possible due to the ‘‘pinching’’ that occurs at the end points.

Another two-parameter weighting function that also varies curvature and elevation separately was proposed by Prelec (1998). The functional form is

$$w(p) = \exp(-\delta(-\log(p))^{\gamma}). \tag{4}$$

For typical values of probability used in most empirical studies (including the present one), it will not be possible to distinguish the linear in log odds function from the Prelec function because both functions can be linearized, and these linearized forms are themselves closely linear for probability range (.01, .99). Instead, specially designed studies testing the key property of the linear in log odds form (see Appendix) and the key property of the Prelec form (compound invariance as defined in Prelec, 1998) will have to be conducted.[4] These functions can also be compared on how they predict responses for extremely small probabilities such as .00001.

Note that unless discriminability and attractiveness empirically covary in just the right way, a one-parameter function for w will fit the data inadequately. We show below that two promising one-parameter functions fail to fit the data of approximately two-thirds the subjects. Further, some simple functional forms for w can be rejected on the basis of existing data, e.g., a power w (as suggested by Luce, Mellers, and Chang, 1993, and by Hey & Orme, 1994) and a linear weighting function $w(p) = (b - a)p + a$, with $0 < a < b < 1$, end points $w(0) = 0$ and $w(1) = 1$ (as suggested by Bell, 1985).

GOALS OF THE PRESENT STUDY

Empirical estimation of the weighting function is not straightforward. In order to assess the shape of the weighting function with standard, nonlinear regression methods (least squares or maximum likelihood) it is necessary to assume functional forms. This creates a problem because the quality of the estimation becomes dependent on the choice of functional forms, and in the case of decision making under risk there are two functional forms that need to be assumed (value and weighting functions) as well as an assumption about the functional that combines the two functions (such as the CPT representation). Although standard residual analysis will permit an examination of the global fit, it is not possible to use residual analysis to assess the fit of each component function separately. Some researchers have attempted to side step this problem by performing sensitivity analyses, either of parameter values, as in Tversky and Fox (1995) and Bernstein, Chapman, Christensen, and Elstein (1997), or of functional forms, as in Chechile and Cooke (1997) and Wu and Gonzalez (1996).

Further, given the psychological motivation for the weighting function, there is reason to expect individual variation in the weighting function. Weighting functions across individuals may differ in curvature and/or elevation. If there is substantial individual variability, then some of the properties that have been observed in previous studies at the aggregate level may not hold at the level of individuals (see Estes, 1956, for a related discussion).

In light of these issues we conducted a new study that permits nonparametric estimation of an individual's value function and weighting function under

[4] The critical property of the Prelec function is subproportionality, and subproportionality produces the common ratio effect. The linear in log odds form is mostly subproportional (throughout the 0,1 interval for various combinations of parameters), and the Prelec function is everywhere subproportional. There is limited empirical evidence about the extent to which subproportionality holds throughout the [0,1] interval.

RDU. A main advantage of this approach is that the nonparametric estimation makes fewer assumptions and therefore stays closer to the data. The nonparametric estimation permits tests of the key features of the weighting function without requiring one to assume the feature in order to perform the estimation. One drawback of the nonparametric procedure is that estimates can be noisier (i.e., the standard parametric assumptions serve to ''smooth'' the data). However, the nonparametric and parametric estimation can be used together; the nonparametric estimation provides a way to assess the deviation of the data from parametric forms, thereby avoiding potential misspecification errors.

The present design provides a large number of observations, permitting assessment at the individual level. Further, probabilities and outcomes are varied in a factorial design, which provides an opportunity to assess how certainty equivalent (CE) changes as function of varying p while holding outcomes constant and varying the outcomes while holding p constant. This design feature improves on that initially used by Tversky and Kahneman (1992).

METHOD

Participants

We report data from 10 participants (5 female).[5] All participants were graduate students in psychology. They were paid $50 for participating in four 1-h sessions. In addition, an incentive compatible elicitation procedure was used (Becker, DeGroot, & Marschak, 1964); participants' certainty equivalents were entered into a subsequent auction. In order to have an adequate assessment of an individual's value and weighting function, we opted for a traditional psychophysical paradigm with many trials (hence relatively few subjects).

Materials

The basic design consisted of 15 two-outcome gambles crossed with 11 levels of probability associated with the maximum outcome.[6] The two outcomes gambles were (in dollars) 25–0, 50–0, 75–0, 100–0, 150–0, 200–0, 400–0, 800–0, 50–25, 75–50, 100–50, 150–50, 150–100, 200–100, and 200–150. Note that all gambles offered nonnegative outcomes, so prospect theory codes all such outcomes as gains. The 11 probability levels were .01, .05, .10, .25, .40, .50, .60, .75, .90, .95, and .99. Nine of these gambles (randomly chosen) were repeated to provide a measure of reliability. Except for the restriction that the identical gamble could not appear in two consecutive trials, the repeated gambles were randomly interspersed within the complete set of gambles.

Procedure

A computer program following the procedure outlined in Tversky and Kahneman (1992) was used in this study. The program presented one gamble on the screen and asked the partici-

[5] Data from an 11th participant were discarded because his responses appeared random and inconsistent, and violated monotonicity.

[6] In addition to the two-outcome gambles, 36 three-outcome gambles were included. Data from these gambles will be presented elsewhere.

pant to choose a certainty equivalent from a menu of possibilities. The format is illustrated below for a gamble offering a 50% chance to win \$100 or \$0 otherwise.

Money (no gamble)	Prefer Sure Thing	Prefer Gamble
100		
80		
60		
40		
20		
0		

The screen for this particular gamble offered the participant a choice between sure amounts of 100, 80, 60, 40, 20, and 0 dollars. For each row in the table, participants checked whether they preferred the sure amount or the gamble. The range of the choices spanned the range of the gamble. To illustrate, consider a participant who preferred a sure amount of \$60 to the gamble, but preferred the gamble to \$40 for sure. The participant would place check marks as follows:

Money (no gamble)	Prefer Sure Thing	Prefer Gamble
100	√	
80	√	
60	√	
40		√
20		√
0		√

From this response we infer that the certainty equivalent for this participant must be somewhere between \$40 and \$60. Once the certainty equivalent was determined within a range (such as between \$40 and \$60 in the above example), a second screen with a new menu of choices was presented using a narrower range (e.g., \$40 to \$60 in increments of \$4). By repeating this process, the program approached the certainty equivalent to the nearest dollar (taking the midpoint of the final range).

The program forced the response checkmarks to be (1) all in the left column indicating a preference for the sure amount, (2) all in the right column indicating a preference for the gamble, or (3) at some row switch from the left column to the right column indicating a switch in preference from the sure amount to the gamble. At most one switch (from left to right) was permitted on each screen. Note that the procedure forced internality of the certainty equivalent response; that is, all CE were forced to be between the highest outcome X_h and the lowest outcome X_l (symbolically, $X_l \leq \mathrm{CE} \leq X_h$). We adopted the convention that on each screen the certainty equivalent options appeared in descending order.

Note that the elicitation procedure did not require participants to generate a certainty equivalent. Rather, participants chose certainty equivalents from a menu (for related procedures see Tversky & Kahneman, 1992, and Bostic, Herrnstein, & Luce, 1990). The program presented gambles in random order to each participant over four sessions lasting approximately 1 h each.

RESULTS

There are many ways to analyze these CE data. The standard approach has been to assume functional forms for v and w and then to perform a goodness of fit test to evaluate those forms for v and w. A major limitation of this approach is that there is no way to independently assess the fit of v and w (e.g., standard residual analysis is on the CE rather than the individual component functions). The approach taken in this paper is to use a nonparametric procedure that does not make assumptions about specific functional forms. We then use these nonparametric estimates to evaluate specific functional forms for v and w. If specific functional forms fit the nonparametric estimates well, then it will be appropriate to perform the more traditional estimation procedure with those functional forms substituted for v and w. Thus, the nonparametric estimation procedure used here provides a method for assessing the functional forms of v and w in the context of a bilinear model such as CPT.

The results section is organized as follows. The first subsection presents descriptive statistics on the aggregate data and examines the reliability of the certainty equivalents. The second subsection describes a nonparametric algorithm for estimating the value function and the weighting function. The third subsection presents the results from that nonparametric algorithm for both aggregate and individual data. The fourth subsection provides parametric fits to the nonparametric estimates directly. The fifth subsection uses the nonparametric estimates to assess the property of subcertainty. Finally, the sixth subsection presents the results of a traditional nonlinear regression, both to provide a comparison to previous studies and to permit a comparison between the standard estimation approach and the new, nonparametric approach presented above.

Median Certainty Equivalents and Reliability

The median certainty equivalent for each of the 165 two-outcome gambles appears in Table 1. Although weak monotonicity is violated in 21% of the pairwise comparisons, there does not appear to be a systematic pattern to these violations.

Reliability for the nine repeated gambles was measured by the maximum likelihood estimator of the intraclass correlation (Haggard, 1958). The intraclass correlation on the median data was .99 (the median absolute deviation was $1.50). The median of the 10 intraclass correlations computed on the individual subject data was .96, with a range of .60 to .99. Only two intraclass correlations were below .90 (the lowest .60 and the second lowest

TABLE 1
Median Certainty Equivalent for Each Gamble ($N = 10$)

	Probability attached to higher outcome										
Outcomes	0.01	0.05	0.1	0.25	0.4	0.5	0.6	0.75	0.9	0.95	0.99
25–0	4.0	4.0	8.0	9.0	10.0	9.5	12.0	11.5	14.5	13.0	19.0
50–0	6.0	7.0	8.0	12.5	10.0	14.0	12.5	19.5	22.5	27.5	40.0
75–0	5.0	10.0	8.5	14.0	17.0	16.0	18.0	23.0	36.0	31.5	48.5
100–0	10.0	10.0	15.0	21.0	19.0	23.0	35.0	31.0	63.0	58.0	84.5
150–0	10.0	10.0	18.0	25.0	34.0	25.0	49.0	70.0	41.5	106.0	118.0
200–0	6.0	9.0	21.0	26.5	34.5	34.0	56.0	48.0	80.0	102.0	158.0
400–0	18.0	24.0	24.0	54.0	64.0	58.0	58.0	115.0	208.5	249.0	277.0
800–0	9.5	42.0	36.0	90.0	91.0	89.5	207.0	197.5	404.0	448.5	519.0
50–25	28.0	29.5	30.5	32.0	34.5	34.5	37.5	36.5	38.0	41.5	41.0
75–50	56.5	58.0	56.0	59.5	62.0	63.0	64.0	64.5	64.5	65.0	68.5
100–50	58.0	59.0	59.0	62.5	66.5	70.0	78.0	82.5	80.0	78.5	89.0
150–50	57.0	58.5	71.0	79.0	89.0	84.0	92.0	99.0	121.0	106.0	117.0
150–100	114.0	110.0	116.5	116.0	121.0	125.0	125.5	131.0	133.5	130.0	142.5
200–100	111.5	115.0	117.0	123.0	131.0	135.0	144.0	146.0	149.0	171.0	158.0
200–150	156.0	165.0	160.0	166.0	171.5	170.5	176.0	177.0	187.5	179.5	190.0

.87), so the procedure elicited relatively high levels of reliability for nine of the 10 participants. Analogous results were observed for other measures of reliability such as the Pearson correlation, Kruskal and Goodman's γ, and Kim's *d*.

Nonparametric Estimation Algorithm

Estimation of the value function and weighting function in the context of utility theory presents challenging problems. A major stumbling block is the need to use the inverse of the value function in estimation. For example, CPT represents the certainty equivalent as

$$v(\mathrm{CE}) = w(p)v(X) + [1 - w(p)]v(Y). \tag{5}$$

In an experiment, however, one observes the CE rather than v(CE). Mathematically, it is easy to apply the inverse of v to both sides of Eq. 5, but the inverse of v is difficult to estimate empirically.

Many researchers have side stepped this problem by assuming a functional form for v and using the known inverse of that v directly in the estimation procedure such that:

$$\mathrm{CE} = v^{-1}\,\{w(p)v(X) + [1 - w(p)]v(Y)\}, \tag{6}$$

where v^{-1} denotes the inverse of v. Examples of this approach appear in Birnbaum and McIntosh (1996), Chechile and Cooke (1997), Luce et al. (1993), and Tversky and Kahneman (1992). Having made an assumption for the functional form of v, previous researchers tested several functional forms for the weighting function w in what can be described as a ''goodness of fit contest'' for different choices of v and w.

As stated above, the standard nonlinear regression technique does not permit an examination of residuals for v and w separately. To circumvent this problem and attempt an estimation with as few assumptions as possible, we developed and implemented a nonparametric algorithm to estimate v and w in the context of CPT without having to specify their functional forms. The algorithm treats the levels of v and w as parameters to estimate (i.e., $v(25)$, $v(100)$, $w(.50)$, etc., are treated as parameters). The parameters are estimated using an alternating least squares approach in which at each iteration either the w's are held fixed while the v's are estimated, or the v's are held fixed while the w's are estimated. The remainder of this section describes the algorithm in more detail. Readers not interested in these details may skip to the next subsection without loss of continuity.

We propose an algorithm for dealing with the bilinear estimation problem. The intuition is based on an alternating least squares algorithm used successfully to solve other scaling problems (e.g., De Leeuw, Young, & Takane, 1976): divide the problem into manageable subparts, estimate parameters for each subpart, and iterate over the subparts until an optimum is found. While general convergence properties of this algorithm will not be discussed here, analyses for all subjects (and median data) converged relatively quickly; the median number of iterations was 6.

We assumed Eq. (5) with an additive, normally distributed error term on the v scale as in

$$v(\mathrm{CE}) = w(p)v(X) + [1 - w(p)]v(Y) + \epsilon_v. \tag{7}$$

The subscript on ϵ is to acknowledge that the error is modeled on the scale v (in contrast to ϵ in a different analysis shown below that is on the CE scale). Using the certainty equivalents from the 165 two-outcome gambles, pick starting values for the 11 $w(\ )$s (i.e., one for each p) and the eight $v(\ )$s (i.e., one for each dollar amount). The algorithm proceeds as follows, with superscript denoting the ith iteration:

1. Interpolate for $v^i(\mathrm{CE})$: using the estimates of $v(\ )$ for the current iteration, which are based on the eight stimuli dollar amounts, interpolate to find $v^i(\mathrm{CE})$ for each of the 165 certainty equivalents; these 165 $v^i(\mathrm{CE})$'s will be used as ''data'' for the estimation in Step 2 and Step 3.

2. Fix all $v(\ )$'s to the current iteration values and estimate the eleven w^i's using an iteratively reweighted, nonlinear least squares algorithm.

3. Fix the 11 $w(\ )$'s to the current iteration values and estimate the eight $v^i(\ )$'s using an iteratively reweighted, nonlinear least squares algorithm.

4. If an optimum is found, then stop; otherwise, increment iteration counter i and repeat.

The algorithm was implemented in the statistical package *Splus* (Becker, Chambers, & Wilks, 1988) and made use of the package's nonlinear least squares algorithm as well as its interpolation algorithm. We used linear interpolation in Step 1. The nonparametric algorithm did not impose monotonicity on v or w, though the restrictions $v(0) = 0$ and $0 < w(p) < 1$ for $0 < p < 1$ (as imposed by the general rank-dependent theory) were implemented into the estimation procedure. Convergence was defined when the change in fit was within 10^{-4}, the fit was the weighted sum of squared error from the nonlinear regression, and starting values were chosen using the functional forms and parameter estimates from the median data given in Tversky and Kahneman (1992) with an estimated exponent of .88 for the power value function $v(x) = x^\alpha$ and .61 for the exponent of the weighting function $w(p) = p^\beta/(p^\beta + (1 - p)^\beta)^{1/\beta}$.[7]

[7] For other possible algorithms in the fledgling area of ''data augmentation procedures'' see McLachlan and Krishnan (1997) or Tanner (1996). For other approaches to the problem of testing and estimating utility theories see Coombs et al. (1967) and Chechile & Cooke (1997). While we incorporated a linear interpolation to estimate v, similar results were observed with a cubic spline. Further, interchanging the roles of w and v in Steps 2 and 3 did not change the solution appreciably nor did different starting values, both providing promising

An iteratively reweighted, nonlinear regression was used in Steps 2 and 3 because the magnitude of the residuals depends, in part, on the size of the outcomes. For instance, the error associated with the gamble (25, .5; 0, .5) is likely to be less than the error associated with the gamble (800, .5; 0, .5) because the latter provides a larger range for the subject's CE.[8] Our data consist of one observation per gamble so we cannot assess nor model intrapersonal variability. Our strategy for dealing with this problem was to examine the residuals for systematic patterns and to examine the standard error of the parameters (i.e., the estimated *w*'s and *v*'s) for homogeneity. An iteratively reweighted, nonlinear regression, where the weights on the residuals depend on the parameter estimates, appeared to work well. For 8 of the 10 subjects, the weights $1/\log(\widehat{\mathrm{CE}}^{i})$, where $\widehat{\mathrm{CE}}^{i}$ denotes the predicted certainty equivalent for that gamble, yielded fits with no apparent systematic pattern in the residuals and relatively uniform standard errors across the parameters; the weights $1/\mathrm{sqrt}(\widehat{\mathrm{CE}}^{i})$ produced better results (no apparent systematic pattern in residuals and relatively uniform errors) for the remaining two subjects (Subjects 2 and 4).[9]

The main advantage of this nonparametric algorithm is that it provides estimates for *v*( ) and *w*( ). We call this algorithm ''nonparametric'' because we do not assume functional forms of *v* or *w*. Instead, the algorithm estimates the values of *v*( ) and *w*( ) at the levels of the stimuli, and the interpolation in Step 1 eliminates the need for an inverse because the estimation occurs on the scale *v*.

Nonparametric Estimation of the Value and Weighting Functions

Figure 5 presents the nonparametric fits of *v* and *w* for the median data presented in Table 1 under the assumption of the CPT representation. The error bars are $\pm$ 1 SE,[10] estimated from the inverse Hessian of the fit functions at Steps 2 and 3, respectively.

---

signs that a global optimum was found. Step 1 served to change the scale of the data at each iteration; this is in the spirit of ''optimal scaling'' techniques (e.g., Young, 1981).

[8] Interpersonal variability supports this intuition. We computed the interquartile range (IQR) over the 10 subjects for each of the 165 gambles. The IQR tend to be monotonically increasing with the greater outcome, and also monotonically increasing in p, which is counter to an intuitive error model that has maximal error near $p = .5$ (in the spirit of the binomial distribution).

[9] We also examined more complicated weighting schemes but they did not improve the fits relative to the weighting described in the text. These additional weighting schemes included allowing the weights (a) to depend on the probabilities of the gamble $p(1 - p)$, (b) to depend on the current w, i.e., $w(1 - w)$, (c) to be based on the expected value of the gamble, and (d) to be based on the log of the expected value. For a discussion of weighting in the context of nonlinear regression see Carroll and Rupert (1988).

[10] The standard error bars are wider for some values of *v*( ) than others. Recall that not all dollar amounts appeared equally often in the stimuli set; e.g., the values 400 and 800 each appeared in 11 gambles (always as the greater outcome), whereas the value 50 appeared in 55 gambles (as the greater outcome in 22 of those 55 gambles).

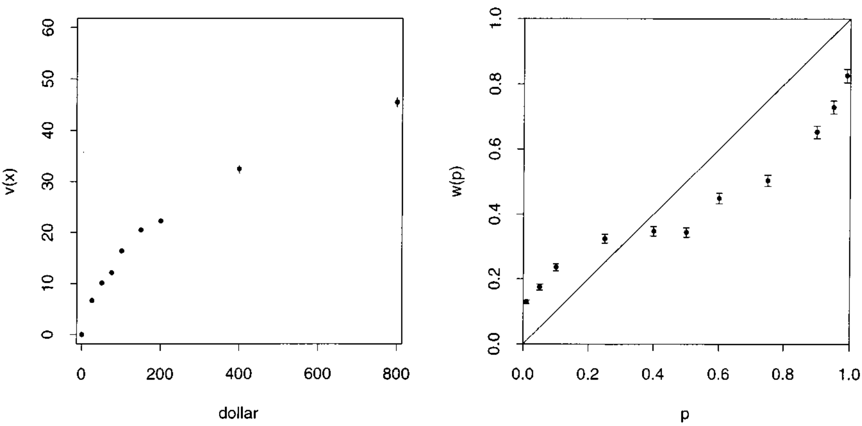

FIG. 5. Estimates of v and w for median data using the nonparametric, alternating least squares algorithm. $N = 10$ with 165 gambles per subject. Error bars are $\pm$ 1 SE, estimated from the inverse Hessian at Steps 2 and 3.

The general characteristics of v and w for the median data are consistent with previous findings, even though the present estimation procedure imposed relatively little structure on the data. The value function v is concave and the weighting function has an inverse-S-shape (concave, then convex) with a crossover point well below $p = .5$ (p between .25 and .40).

The same algorithm was applied to individual subject data. As shown in Fig. 6, the individual participants each exhibited a pattern similar to the median data. That is, the concavity of v and the inverse S-shape of w appear to be a regularity that holds both at the aggregate level and for individual subjects, with the weighting function for Subject 6 appearing to be concave everywhere. The standard errors, though approximate, provide an index of the quality of the fit. Subject 8 was the participant with low reliability.

Note that there is substantial heterogeneity in curvature (discriminability) and in elevation (attractiveness) of the weighting function, and the two effects appear to be somewhat independent. For example, Subject 6 predominantly overweights p (relative to the identity line) whereas Subject 9 predominantly underweights p; both participants exhibit roughly the same degree of curvature. Subjects 1 and 7 appear to cross the identity line at roughly the same level, yet they exhibit different degrees of curvature. Subject 1 discriminates p (though is still regressive) in the range [.1 and .9], whereas Subject 7 tends to weight the probabilities in that range at the same level. Given that curvature and elevation appear to vary somewhat independently across participants, it would be surprising if a one-parameter function could fit all 10 subjects. Indeed, in the next subsection we show that although two promising one-parameter functions fail to capture the pattern of 6 of the 10

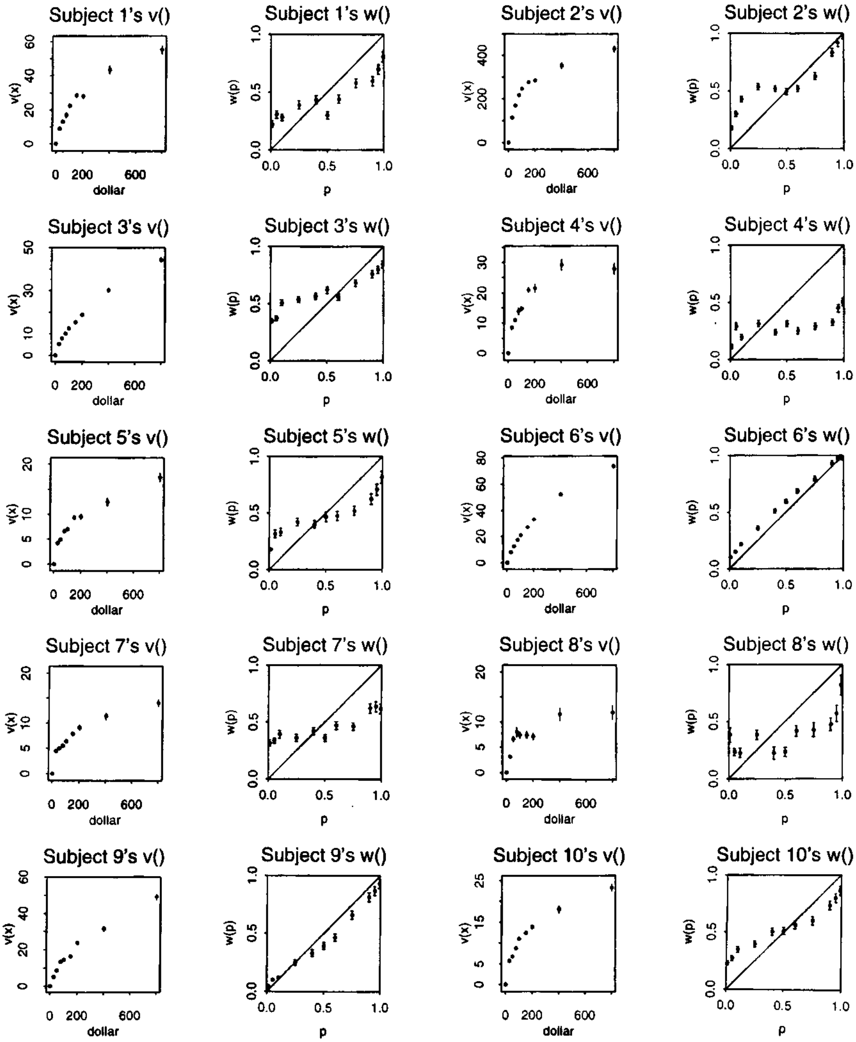

FIG. 6. Estimates of *v* and *w* for each individual participant using the nonparametric, alternating least squares algorithm. 165 gambles per subject. Error bars are ± 1 SE, estimated from the inverse Hessian at Steps 2 and 3. Note that value plots are scaled to each participant's own *v*.

participants, the two-parameter, linear in log odds form cannot be rejected for any of the 10 subjects.

We now turn to an interpersonal assessment of attractiveness and discriminability on the estimates of *w*( ) from the individual subject data. We performed a sign test on all pairs of subjects to test for differences in elevation.

Let $w_i(p)$ be the estimated $w(p)$ for Subject i. To test for differences in elevation, for all pairs of subjects i, j (where $i \neq j$), we counted the number of p levels in which $w_i(p) > w_j(p)$. Each pair of subjects provided 11 such comparisons. Sixteen of the 45 subject pair comparisons yielded elevation differences that were statistically significant at $\alpha = .05$ (i.e., 10 or more of the 11 comparisons by the two-tailed binomial test) and an additional 6 subject pairs were significant at $\alpha = .064$ (which corresponds to 9 or more of the 11 comparisons by the two-tailed binomial test).

An analogous test was performed for differences in discriminability. Differences in curvature between subject pairs were at the level of chance. The test for discriminability involved comparing a within-subject difference for Subject i (e.g., $w_i(.05) - w_i(.01)$) with a within-subject difference for Subject j for the same two stimulus levels (e.g., $w_j(.05) - w_j(.01)$). The size of the standard errors of the estimates and the lack of power in the binomial test are the main reasons that discriminability differences were not detected with the nonparametric estimates, even though elevation differences were detected. We show below that a parametric test of discriminability is sufficiently powerful to detect interpersonal differences.

These nonparametric estimates for v and w were based on the assumption of a rank-dependent representation or combination rule. That is, the estimation algorithm assumed that functions v and w were combined using Eq. (2). This assumption can be evaluated by fitting the more general model

$$w(p)v(X) + g(1 - p)v(Y) \tag{8}$$

which relaxes the restriction on the weighting function but maintains the bilinear form (see Miyamoto, 1988). We estimated the more general model and found that $g(1 - p)$ of Eq. (8) roughly coincides with $1 - w(p)$ of CPT (Eq. (2)). The intraclass correlation, which serves as an index of the fit to the identity line, between $g(1 - p)$ and $1 - w(p)$ for the median data was .99; the median intraclass correlation over the 10 participants was .96. Similarly, the $w(p)$'s from Eq. (2) and (8) also coincided: an intraclass correlation of .99 on both the median data and the median participant. Thus, functions w and g in Eq. (8) appear to sum to 1 as required by CPT.

Nonparametric Assessment of the Linear in Log Odds w

In the introduction we argued for a two-parameter probability weighting function (Eq. (3)), where the parameters correspond to the psychological properties of attractiveness and discriminability. One way to examine this functional form (as well as others) is to fit the linear in log odds function directly to the nonparametric estimates of $w(\,)$. We also fitted a power value function to the nonparametric estimates of $v(\,)$, i.e., $v(X) = \theta X^{\alpha}$, where the extra parameter θ accounts for the arbitrary scale of the nonparametric v and carries no psychological meaning in this context.

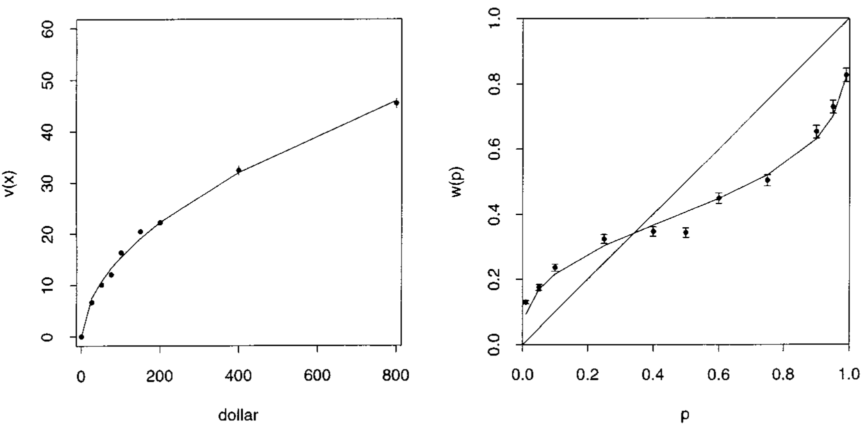

FIG. 7. Same as Fig. 5 with best fitting power function for v and best fitting linear in log odds w overlays.

Figures 7 and 8 show the best fitting power v and linear in log odds w superimposed on the nonparametric estimates for the median data and individual subject data. The data are fit quite well with a power value function, i.e., $v(X) = \theta X^{\alpha}$ (9 of 10 participants) and the two-parameter linear in log odds w parameter weighting function (all participants).

A runs test provides a formal test of how well a specific functional form fits the nonparametric estimates. The logic behind this test is that a poor fitting function would produce residuals that have more (or fewer) runs than expected by a simple chance model where the signs of the residuals are independent. To test the linear in log odds weighting function, first observe that the left-most point in the right panel of Fig. 7 (i.e., the estimated $w(.01)$ for the median data) is above the curve. Moving left to right, the second, third, and fourth points are also above the curve; the fifth point is below the curve, etc., yielding a total of four runs, which fails to reject the null hypothesis of the runs test (Siegel, 1956). In this runs test, the residuals are categorized as being above or below the nonlinear regression curve and the pattern is compared to a chance model. The null hypothesis for the runs test was not rejected for the linear in log odds function for the median data or any of the 10 individual subject weighting functions. The runs test yielded comparable results for the two-parameter Prelec function (Eq. (4)) on the individual data (except that data for Subject 7 was rejected by the runs test) and the median data. Although it may appear that the runs test is not sufficiently powerful, the one-parameter special case of the Prelec function (Eq. (4) with $\delta = 1$) and the one-parameter function suggested by Tversky and Kahneman (1992) were each rejected by the runs test for 6 of the 10 participants. The runs test failed to reject either of the one-parameter functional forms for the median data.

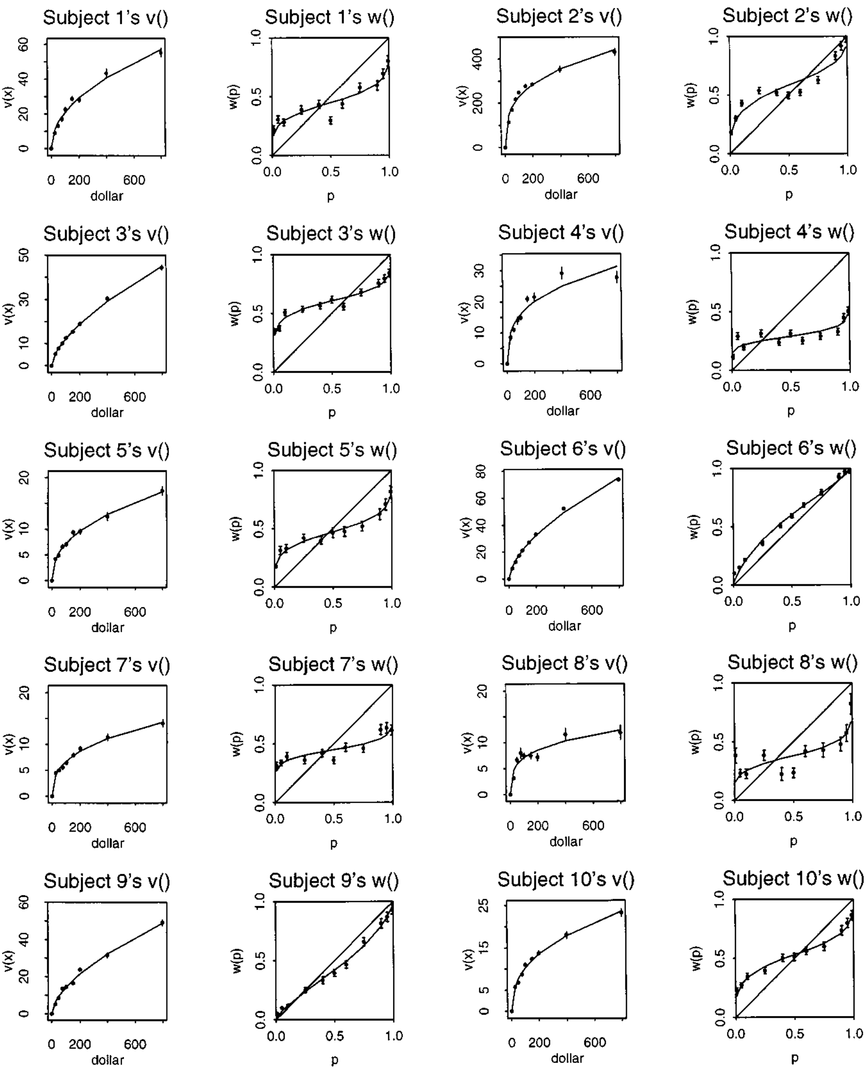

FIG. 8. Same as Fig. 6 with best fitting power function for *v* and best fitting linear in log odds *w* overlays. Note that value plots are scaled to each participant's own *v*.

We next turn to a more standard method of assessing goodness of fit for *w*. Recall that a feature of Eq. (3) is that it is linear in the log odds scale. A standard measure of fit to linearity is the R^2 from the linear regression. We assessed the fit of Eq. (3) by regressing *w* in log odds units on probability in log odds units. The R^2 for the median data was .98 and the median R^2 over the 10 subjects was .92. Subject 8 (the subject who exhibited the lowest

TABLE 2
Test of Subcertainty Using the Nonparametric Estimates *w*( )

	Probability pairs						
Subject	(.01, .99)	(.05, .95)	(.1, .9)	(.25, .75)	(.4, .6)	(.5, .5)	Binomial test
1	1.03	0.99	0.87	0.96	0.87	0.59	Trend sub
2	1.16	1.22	1.26	1.16	1.04	0.99	Trend super
3	1.20	1.17	1.27	1.22	1.12	1.24	Super
4	0.53	0.71	0.45	0.46	0.40	0.59	Sub
5	0.99	1.02	0.96	0.94	0.88	0.93	Trend sub
6	1.08	1.12	1.15	1.15	1.20	1.19	Super
7	0.93	0.97	1.01	0.82	0.89	0.72	Trend sub
8	1.20	0.81	0.70	0.81	0.64	0.47	Trend sub
9	0.98	0.97	0.93	0.91	0.80	0.79	Sub
10	1.09	1.07	1.09	0.99	1.07	1.03	Trend super
Median data	0.97	0.94	0.91	0.90	0.86	0.87	Sub

reliability) had the lowest R^2 of .56. The two-parameter Prelec function performed slightly better by this measure: the R^2 for the median data was .99, the median R^2 over the 10 subjects was .95, and the lowest R^2 (Subject 8) was .75.

Subcertainty

Recall that subcertainty is defined as $w(p) + w(1 - p) < 1$, and it can be interpreted as a measure of probabilistic risk aversion (see Wakker, 1994). When $w(p) + w(1 - p)$ is substantially less than 1, the individual tends to avoid risk, whereas when $w(p) + w(1 - p)$ exceeds 1, the individual tends to embrace risk. Subcertainty is related to the elevation of w or how attractive the decision maker regards gambling.

The nonparametric estimates of $w(\ )$ reported above can be used to examine subcertainty with the six probability pairs that sum to 1, i.e., (.01, .99), (.05, .95), (.1, .9), (.25, .75), (.4, .6), and (.5, .5). Table 2 presents the results of the binomial test on the null hypothesis that for a given subject (or aggregate data) the sum of each of the six w pairs = 1. Because each participant contributed six tests of subcertainty we do not have a very powerful test—if all six sums are on the same side of 1, the null hypothesis can be rejected at two-tailed $\alpha = .032$. Subjects 4 and 9 (as well as the median data) had perfect patterns of subcertainty. Two subjects (3 and 6) had perfect supercertainty patterns (i.e., sum greater than 1). Subjects 1, 5, 7, and 8 produced five of six sums in the direction of subcertainty, and Subjects 2 and 10 produced five of six sums in the direction of supercertainty (by the binomial test, five of six corresponds to a two-tailed α of .218). There did not appear to be a pattern to the sole violation. In sum, even though the median data exhibited subcertainty quite clearly (consistent with previous studies), 4 of the 10 participants had patterns suggesting a weighting function consistent with the property of supercertainty. It is in this sense that the aggregate data may not paint a picture that is consistent with all data at the individual subject level.

Related properties of lower subadditivity (LS) and upper subadditivity (US) were discussed by Tversky and Wakker (1995). LS is $w(q) \geq w(p + q) - w(p)$ for $p + q$ bounded away from 1, i.e., increments from $p = 0$ have larger weights than identical increments in the interior of [0, 1], and US is $1 - w(1 - q) \geq w(p + q) - w(p)$, i.e., increments from $p = 1$ have larger weights than identical increments in the interior. If both LS and US hold, then w is said to exhibit bounded subadditivity. The present design allowed for three tests of LS using the nonparametric estimates of w (stimuli probability levels of .05 vs. .10, .25 vs. .5, and the triple .25, .5, and .75). Only one violation of LS was observed—the triple for Subject 9. Similarly, three tests of US were conducted (stimuli probability levels of .95 vs. .90, .75 vs. .50, and the triple .75, .50, and .25). Subject 6 violated US on all three comparisons; Subject 2 violated US on the .95 vs. .90 comparison. Thus, 7 of 10 subjects exhibited no violations of bounded subadditivity, 2 subjects exhibited a trend for bounded subadditivity (i.e., one inconsistent pattern out of a possible six), and 1 subject violated US.

Standard Parametric Fits of the Weighting Function and Value Function

Given the nonparametric estimates of the value function and weighting function presented above, we can now proceed with the more standard nonlinear regression technique. The main difficulty with the standard approach is that functional forms for v and w need to be assumed in order to perform the estimation (and in the case of CPT, the inverse of v needs to be applied). Without independent estimates of v and w it is difficult to evaluate the fits of the functions apart from global indices such as root mean square error. The nonparametric fits presented above provide a means for assessing the appropriateness of functional forms prior to using those functional forms in parametric estimation. As shown above, the power function v and the linear in log odds w fit the individual subject data quite well.

The results of the standard nonlinear least squares estimation on the individual subject data appear in Fig. 9. The parameter estimates from this nonlinear least squares regression are presented in Table 3. For the median data, the α estimate for the power value function $v(X) = X^{\alpha}$ (the parameter θ was not included here) was .49, and the estimates of δ and γ for the linear in log odds w were .77 and .44, respectively.

By comparing Figs. 6 and 9, one can see that the fits of this parametric analysis resemble the nonparametric results. Thus, there is convergence between the two techniques, one technique that assumed relatively little structure and the other that assumed specific functional forms. This nonlinear regression did not weight residuals as was done in the nonparametric estimation; thus under the assumption of normally distributed error, these nonlinear

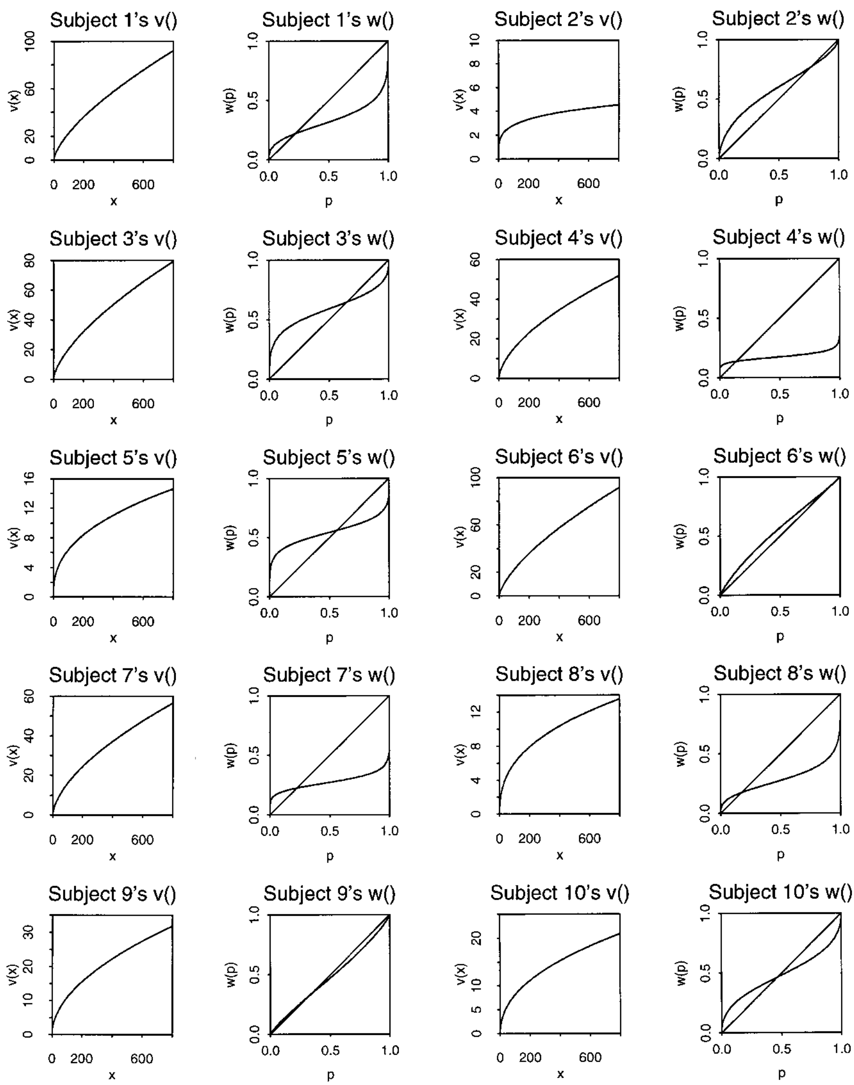

FIG. 9. Estimates of power *v* and linear in log odds *w* for each individual participant using a standard nonlinear algorithm. 165 gambles per subject. Note that value plots are scaled to each participant's own *v*.

TABLE 3
Parameter Estimates from the Standard Nonlinear Least Squares Regression with a Power Value Function (Exponent = α) and the Linear in Log Odds Weighting Function (Eq. (3))

	Parameters				
Subject	α	δ	γ	PRE	PRE-EU
1	0.68 (.10)	0.46 (.11)	0.39 (.03)	0.91	0.76
2	0.23 (.06)	1.51 (.46)	0.65 (.04)	0.90	0.87
3	0.65 (.12)	1.45 (.35)	0.39 (.02)	0.95	0.90
4	0.59 (.05)	0.21 (.04)	0.15 (.02)	0.93	0.44
5	0.40 (.08)	1.19 (.32)	0.27 (.02)	0.84	0.80
6	0.68 (.06)	1.33 (.15)	0.89 (.03)	0.99	0.92
7	0.60 (.06)	0.38 (.07)	0.20 (.02)	0.91	0.56
8	0.39 (.07)	0.38 (.11)	0.37 (.04)	0.83	0.65
9	0.52 (.08)	0.90 (.18)	0.86 (.04)	0.98	0.90
10	0.45 (.09)	0.93 (.26)	0.50 (.03)	0.91	0.85
Median data	0.49 (.04)	0.77 (.10)	0.44 (.01)	0.97	0.80

Note. Values in parentheses are standard errors. PRE represents the proportion reduction in (sum of squared) error relative to the simple grand mean model. PRE-EU represents the proportion reduction in error of expected utility with a power value function relative to the grand mean; PRE-EU provides a baseline in which to compare PRE. Thus, the comparison of PRE and PRE-EU shows the contribution of estimating the extra two weighting function parameters relative to fixing both parameters to 1.

least squares estimates are the maximum likelihood estimates[11] (Seber & Wild, 1989). The error for this estimation procedure was on the CE scale. Table 3 also presents the proportion reduction in sum of squared error relative to the simple grand mean model and the proportion reduction in error of expected utility with a power value function relative to the grand mean model.

We constructed Z tests based on the parameter estimates and standard errors in Table 3 to assess interpersonal differences in discriminability and attractiveness. For all pairs of subjects i, j (where $i \neq j$), we computed the Z test between δ_i and δ_j, and also the difference between γ_i and γ_j. Forty of the 45 interpersonal Z tests comparing γ (the discriminability parameter) between participants were statistically significant at $\alpha = .05$ using a pooled standard error. Twenty-five of the 45 interpersonal Z tests comparing δ (the

[11] There was high multicolinearity between the δ parameter of the linear in log odds w and the exponent of v (as measured by the off-diagonal elements of the inverse Hessian matrix). For the median data, these two parameters correlated $-.98$ (as was the median correlation between these two parameters over the 10 participants). The curvature parameter for w and the exponent of v were moderately correlated (.56 for the median data) as was the correlation between the two parameters of the linear in log odds w ($-.66$ for the median data). The two-parameter Prelec function also exhibited high multicolinearity between the δ parameter and the α parameter (.97 for both the median data and the median subject).

attractiveness parameter) were statistically significant. Thus, the test on the discriminability parameters was significantly more powerful than the test based on the nonparametric estimates that was presented above (but the test was comparable for attractiveness).

DISCUSSION

A large family of descriptive theories of decision making under risk propose three functions: a transformation of outcomes, a transformation of probabilities, and a functional that combines the transformed outcomes with the transformed probabilities. One reason this classification is attractive is because the widely accepted normative theory, expected utility, is a special case. Under expected utility, outcomes are transformed by the utility function *u*, the transformation on the probabilities is the identity function, and the combination rule is the sum of the utilities weighted by their associated probabilities.

Another member of this family, cumulative prospect theory, has emerged as perhaps the most promising descriptive model of risky decision making (Camerer, 1992, 1995). Under CPT, outcomes are transformed by an S-shaped value function *v*, which is concave for gains, convex for losses, and steeper for losses than gains; probabilities are transformed by an inverse-S-shaped probability weighting function *w;* and values and probability weights are combined through a rank-dependent weighting scheme.

Prospect theory's value function *v* has been widely studied and applied to a number of areas including economics, finance, marketing, medical decision making, organizational behavior, and public policy (for examples see Camerer, 1995, as well as Heath, Larrick, & Wu, 1999). The probability weighting function, in contrast, has received less empirical and theoretical attention. For a few applications of *w* in medical decision making and insurance see Wakker and Deneffe (1996), Wakker and Stiggelbout (1995), Schlee (1995), and Wakker, Thaler, and Tversky (1997). The probability weighting function may offer a rich source of ideas for researchers interested in studying probabilistic risk aversion.

Empirical regularities on the probability weighting function (such as inverse-S-shaped) have emerged that organize choice behavior quite well (see, for example, Camerer, 1995), and psychological principles that can account for the shape of the weighting function have been formulated and tested. For example, the concept of diminishing sensitivity unifies under one explanatory umbrella the curvature of the weighting function, the curvature of the value function, and the subadditivity of probability judgments (Tversky & Kahneman, 1992; Tversky & Koehler, 1994).

We see this paper as making three contributions to the psychological understanding of the probability weighting function: methodological, empirical, and theoretical. First, a new nonparametric algorithm for estimating the weighting function was developed and implemented. With standard estimation techniques, the choice of parametric forms for *w* and *v* often reduces to a problem of global goodness of fit where component functions cannot be evaluated independently. However, the nonparametric algorithm permits an evaluation of component functions without having to assume specific functional forms. In this sense the data were allowed ‘‘to speak for themselves.’’ As we showed under Results, when the nonparametric estimates are available, specific functional forms can be evaluated directly against the nonparametric estimates.

Second, empirical regularities about the weighting function both at the level of aggregate behavior and at the level of individual subjects were identified. The results for the median data add to a growing body of evidence that concave value and inverse-S-shaped weighting functions appear to fit median data quite well (e.g., Abdellaoui, 1998; Camerer & Ho, 1994; Tversky & Kahneman, 1992; Wu & Gonzalez, 1996, 1998). Moreover, our analyses showed that a one-parameter weighting function and a one-parameter value function (power) provide an excellent, parsimonious fit to the median data (see also Camerer & Ho, 1994; Tversky & Kahneman, 1992). In particular, one-parameter weighting functions proposed by Tversky and Kahneman (1992) and by Prelec (1998) fit the median data almost as well as the two-parameter, linear in log odds weighting function.

At the level of individual subjects, however, the story is more complicated. There are two main findings about the weighting function for individual subjects, one about regularity and one about heterogeneity. The regularity over the 10 subjects is that the weighting function, estimated from an individual's data, is inverse-S-shaped, consistent with the psychological property of diminishing sensitivity. It is noteworthy that the inverse-S-shape of the weighting function, the only regularity at the level of individual subjects found in this study, is the main principle needed to explain the examples and survey questions given in the introduction.

In contrast to the regularity of the inverse-S-shaped weighting function, there is a striking amount of heterogeneity in the individual weighting functions, in terms of both the degree of curvature and the elevation. Consistent with this heterogeneity, subcertainty, a measure of probabilistic risk aversion which holds quite well at the level of the aggregate data, failed to hold for 4 of the 10 participants.

The third contribution of this paper is to provide empirical support for a psychological interpretation of the probability weighting function. We argued that the constructs of discriminability and attractiveness could be operationalized in terms of the degree of curvature of the weighting function and elevation, respectively (see also Tversky & Wakker, 1995). These properties were observed to be somewhat independent across the 10 participants. Further, we suggested a two-parameter functional form, the linear in log odds function, that (essentially) models these two constructs independently. Note, however, that the present data could not discriminate between the linear in log odds w and the two parameter weighting function suggested by Prelec (1998).

The present study compared discriminability and attractiveness interpersonally and found that the 10 subjects showed considerable variation on these two dimensions. Such individual variation is hardly striking, particularly if we contrast skilled bridge or poker players to Piaget's children (discriminability) or oil wildcatters with individuals who invest exclusively in government bonds or other low-risk investment instruments (elevation). Many decision theorists would not be surprised at a finding showing individual differences in the utility function. The present data suggest that decision theorists may also want to consider individual differences in the weighting function.

It is interesting to note that these concepts can also vary *intra*personally. The same person may have more knowledge about one domain than another, and hence the weighting function in the latter domain will approach a step function relative to the former. For example, someone with expertise in judging the outcome of a jury trial may be able to discriminate subtle differences in the chances of winning a jury trial, whereas the same person may not have much expertise about football and would not be able to discriminate subtle differences in gambles based on point spreads (see Heath & Tversky, 1991, and Fox & Tversky, 1995, for a related point). An analogous intrapersonal comparison can be made for attractiveness. A person may find one chance domain more attractive than another, as in the comparison between selecting one's own lottery ticket rather than being assigned ticket numbers at random (Langer, 1975).

Kilka and Weber (1998) recently presented evidence supporting these intrapersonal comparisons in the context of uncertainty. Participants were students in a finance course. They priced lotteries based on price changes of either a familiar or an unfamiliar stock. Assuming a power value function, Kilka and Weber found that the δ and γ parameters of the linear in log odds form were greater for the familiar stock. Thus, the decisions based on price changes of the familiar stock exhibited both greater discrimination and greater attractiveness.

However, intrapersonal comparisons on the weighting function that span different sources of uncertainty raise an important question. How can we distinguish poor sensitivity in probability judgment from a flat probability weighting function? This distinction is not relevant in the current paper because probabilities (rather than events) were given to the participants. But, if we want to have a complete understanding of how people make decisions under uncertainty we will need to continue extending our knowledge of how probabilities are estimated and how these estimated probabilities are distorted in decision making. Preliminary attempts to model this distinction appear in Fox and Tversky (1998), Tversky and Fox (1995), Tversky and Wakker (1998), and Wu and Gonzalez (1999).

The Kilka and Weber findings also raise another important point: even though discriminability and attractiveness are logically independent (and can be modelled separately through a two-parameter weighting function), in the real world the two concepts most likely covary across contexts. Attractiveness will tend to be high for domains in which we can make fine discriminations; discrimination will tend to be low in domains that we do not find very attractive. Put simply, we like what we know and we know what we like. If it is true that the weighting function reflects two psychological constructs, then it should be possible to manipulate attractiveness and discrimination independently and to observe corresponding changes in δ and γ, respectively. Experiments designed to manipulate discriminability and attractiveness independently in the spirit of Kilka and Weber are under way. Thus, beyond modeling how decision makers choose among gambles, the probability weighting function provides a foundation for the more general investigation of risk taking, risk attitudes, and decision making under risk and uncertainty.

APPENDIX

Preference Condition for the Linear in Log Odds Weighting Function (Eq. (3))

For two-outcome gambles and the CPT (or rank-dependent) representation, the following preference condition is necessary and sufficient for the linear in log odds weighting function. If the following implication holds for p and q, if

$$(X, p; Y, 1 - p) \sim (X', p; Y', 1 - p)$$

and (A1)

$$(X, p; Y', 1 - p) \sim (X', p; Y'', 1 - p)$$

then

$$(X, q; Y, 1 - q) \sim (X', q; Y'', 1 - q) \tag{A2}$$

with $X > X' > 0$ and $0 < Y < Y' < Y''$, then it holds for all p replaced with $tp/(1 - p + tp)$ and q replaced with $tq/(1 - p + tq)$ with $t > 0$. We call this condition the *linear in log odds condition.* An analogous condition can be written for the case where all outcomes are losses. The need for specifying the condition in terms of $tp/(1 - p + tp)$ will become apparent below; note that on the odds scale, $tp/(1 - p + tp)$ becomes $tp/(1 - p)$. This condition can be generalized to gambles having n outcomes.

The intuition underlying this preference condition is seen by taking the ratio of the odds of p and q; i.e., $p(1 - q)/((1 - p)q)$. The two antecedents provide constraints on p and the conclusion implies a constraint on q. The linear in log odds preference condition is satisfied when the constraints that hold for p and q also hold for $tp/(1 - p + tp)$ and $tq/(1 - p + tq)$, respectively. Note that when the latter two probabilities are converted to odds, the scalar t cancels from the odds ratio. Thus, the odds ratio of each pair of probabilities are identical.

THEOREM 1. For two-outcome gambles (where both outcomes are on the same side of the reference point) CPT implies that the linear in log odds preference condition is necessary and sufficient for the linear in log odds weighting function.

Proof. We make use of the functional equation

$$f(x) + f(y) = f(z)$$

iff (A3)

$$f(tx) + f(ty) = f(tz).$$

If this functional equation holds for all $t > 0$, then under some mild continuity conditions, the only nontrivial solution for monotonic f is

$$f(x) = ax^b \tag{A4}$$

(Aczel, 1966). For a discussion of conditions needed for the continuity of w in the context of rank-dependent models see Wakker (1994).

Our method of proof is to use CPT with outcomes restricted to one side of the reference point to represent the gambles and use Eq. (A3) to provide the necessary constraints on the weighting function. The sketch of the proof will be informal. We first show how CPT and the preference condition together imply the functional equation. Apply CPT to the two antecedents of the preference condition and then add the representations together to yield

$$2\frac{w(p)}{1 - w(p)} = \frac{v(Y'') - v(Y)}{v(X) - v(X')}. \tag{A5}$$

The third indifference (i.e., involving probability q in line 2) yields

$$\frac{w(q)}{1 - w(q)} = \frac{v(Y'') - v(Y)}{v(X) - v(X')}. \tag{A6}$$

Equations (A5) and (A6) imply

$$2\frac{w(p)}{1 - w(p)} = \frac{w(q)}{1 - w(q)}. \tag{A7}$$

Essentially, the ratio of value differences provides a standard sequence that allows w's in odds form to be equated. Apply the same argument to the three indifferences where all probabilities p are replaced with $tp/(1 - p + tp)$ and similarly for all probabilities q to yield the analog of Eq. (A7)

$$2\frac{w(tp/(1 - p + tp))}{1 - w(tp/(1 - p + tp))} = \frac{w(tq/(1 - q + tq))}{1 - w(tq/(1 - q + tq))}.$$

The preference condition requires that once a standard sequence is found for p and q, then standard sequences can also be found for probabilities transformed by $tr/(1 - r + tr)$ where r takes on the value p or q.

Next, define $f(p/(1 - p)) = w(p)/(1 - w(p))$, where f can be conceptualized as a composition of a function that maps odds into $w(p)$ and a function that puts $w(p)$ into odds form. If the preference condition holds for all $t > 0$, then we have a special case of the above functional equation (i.e., assuming that $x = y$ in Eq. (A3)) where the only solution is $f(x) = ax^b$. Substituting back the definition of f into the solution of the functional equation, we have

$$\frac{w(p)}{1 - w(p)} = a \left(\frac{p}{1 - p} \right)^b \tag{A8}$$

and solving for $w(p)$ gives the linear in log odds form of the weighting function (Eq. (A3)). The proof in the other direction follows the analogous argument.

A related procedure that constructs standard sequences in the sense of Eq. (A7) was proposed by Wakker and Deneffe (1996), who were interested in scaling the utility function independent of the weighting function (see also Abdellaoui, 1998). It is interesting that our procedure achieves the opposite goal of specifying a functional form on w independent of the utility function.

On the outcomes, this preference condition resembles the Thomsen condition used in additive conjoint measurement (Krantz, Luce, Suppes, & Tversky, 1971). Loosely speaking, the present condition can be interpreted as the Thomsen condition applied to outcomes with an additional restriction imposed on probabilities. The preference condition for the linear in log odds weighting function can be contrasted with the preference condition given by Prelec (1998) for the two-parameter weighting function he proposed. Unfortunately, both our preference condition and Prelec's compound invariance condition are difficult to test empirically because they require several antecedent conditions. In the context of the CPT axioms, it may be possible to derive consequences of these preference conditions that may be more amenable to empirical test.

REFERENCES

Abdellaoui, M. (1998). *Parameter-free elicitation of utilities and probability weighting functions.* ENS de Cachan, Department of Economics and Management, working paper.

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

Arrow, K. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica,* **19,** 404–437.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science,* **9,** 226–232.

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). *The new S language.* Pacific Grove, CA: Wadsworth.

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations Research,* **33,** 1–27.

Bernstein, L. M., Chapman, G., Christensen, C., & Elstein, A. S. (1997). Models of choice between multioutcome lotteries. *Journal of Behavioral Decision Making,* **10,** 93–115.

Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73–100). Mahwah, NJ: Erlbaum.

Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gamles. *Organizational Behavior and Human Decision Processes,* **67,** 91–110.

Bostic, R., Herrnstein, R. J., & Luce, R. D. (1990). The effect on the preference reversal phenomenon of using choice indifferences. *Journal of Economic Behavior and Organization,* **13,** 193–212.

Camerer, C. (1995). Individual decision making. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 587–703). Princeton, NJ: Princeton Univ. Press.

Camerer, C. F. (1992). Recent tests of generalizations of expected utility theory. In W. Edwards (Ed.), *Utility theories: Measurements and applications* (pp. 207–251). Boston: Kluwer.

Camerer, C. F., & Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty,* **8,** 167–196.

Carroll, R. J., & Ruppert, D. (1988). *Transformations and weighting in regression.* New York: Wiley.

Chechile, R. A., & Cooke, A. D. (1997). An experimental test of a general class of utility models: Evidence for context dependency. *Journal of Risk and Uncertainty,* **14,** 75–93.

Coombs, C. H., Bezembinder, T. G., & Goode, F. M. (1967). Testing expectation theories of decision making without measuring utility or subjective probability. *Journal of Mathematical Psychology,* **4,** 72–103.

De Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika,* **41,** 471–503.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin,* **51,** 380–417.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin,* **53,** 134–140.

Fishburn, P. (1978). On Handa's ''new theory of cardinal utility'' and the maximization of expected return. *Journal of Polical Economy,* **86,** 321–324.

Fishburn, P., & Kochenberger, G. (1979). Two-piece Von Neumann–Morgenstern utility functions. *Decision Sciences,* **10,** 503–518.

Fox, C., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics,* **110,** 585–603.

Fox, C., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science,* **44,** 879–895.

Fox, C. R., Rogers, B. A., & Tversky, A. (1996). Options traders exhibit subadditive decision weights. *Journal of Risk and Uncertainty,* **13,** 5–17.

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

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

Haggard, E. A. (1958). *Intraclass correlation and the analysis of variance.* New York: Dryden.

Hartinger, A. (1998). Do subjective weighting functions of generalized expected value theories capture persisting properties of individual decision makers. Universitat Regensburg, Germany, working paper.

Heath, C., Larrick, R. P., & Wu, G. (1999). *Goals as reference points.* Unpublished manuscript, Duke University and University of Chicago.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty,* **4,** 5–28.

Hey, J. D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica,* **62,** 1291–1326.

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

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

Karmarkar, U. (1979). Subjectively weighted utility and the Allais paradox. *Organizational Behavior and Human Performance,* **24,** 67–72.

Kilka, M., & Weber, M. (1998). *What determines the shape of the probability weighting function under uncertainty?* Universitat Mannheim working paper.

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

Langer, E. (1975). The illusion of control. *Journal of Personality and Social Psychology,* **32,** 311–328.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior and Organization,* **17,** 377–400.

Leland, J. (1998). Similarity judgments in choice under uncertainty: A reinterpretation of the prediction of regret theory. *Management Science,* **44,** 659–672.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 255–295). New York: Academic Press.

Lopes, L. L. (1990). Re-modeling risk aversion: A comparison of Bernoullian and rank dependent value approaches. In *Acting under uncertainty: Multidiscipline conceptions* (pp. 267–299). Boston: Kluwer.

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty,* **4,** 29–59.

Luce, R. D., & Fishburn, P. C. (1995). A note on deriving rank-dependent utility using additive joint receipts. *Journal of Risk and Uncertainty,* **11,** 5–16.

Luce, R. D., Mellers, B. A., & Chang, S. (1993). Is choice the correct primitive? On using certainty equivalents and reference levels to predict choices among gambles. *Journal of Risk and Uncertainty,* **6,** 115–143.

McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions.* New York: Wiley.

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

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy,* **5,** 371–404.

Parker, S., & Schneider, B. (1988). Conjoint scaling of the utility of money using paired comparisons. *Social Science Research,* **17,** 277–286.

Piaget, J. P., & Inhelder, B. (1975). *The origin of chance in children.* New York: Norton.

Prelec, D. (1998). The probability weighting function. *Econometrica,* **66,** 497–527.

Preston, M., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology,* **61,** 183–193.

Quattrone, G. A., & Tversky, A. (1988). Contrasting rational and psychological analyses of political choice. *American Political Science Review,* **82,** 719–736.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization,* **3,** 323–343.

Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model.* Boston: Kluwer.

Schlee, E. E. (1995). The comparative statics of deductible insurance in expected and non-expected utility theories. *Geneva Papers on Risk and Insurance Theory,* **20,** 57–72.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression.* New York: Wiley.

Segal, U. (1989). Anticipated utility: A measure representation approach. *Annals of Operations Research,* **19,** 359–374.

Siegel, S. (1956). A method for obtaining an ordered metric scale. *Psychometrika,* **21,** 207–216.

Starmer, C., & Sugden, R. (1989). Violations of the independence axiom in common ratio problems: An experimental test of some competing hypotheses. *Annals of Operations Research,* **19,** 79–102.

Tanner, M. A. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions.* New York: Springer Verlag.

Thaler, R. H., & Ziemba, W. T. (1988). Parimutuel betting markets: Racetracks and lotteries. *Journal of Econometric Perspectives,* **2,** 161–174.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review,* **102,** 269–283.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business,* **59,** S251–S278.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty,* **5,** 297–323.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review,* **101,** 547–567.

Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica,* **63,** 1255–1280.

Tversky, A., & Wakker, P. (1998). *On the composition of risk preference and belief.* Unpublished manuscript, Stanford University and Leiden University.

Wakker, P. (1989). *Additive representations of preferences.* Dordrecht: Kluwer.

Wakker, P. (1994). Separating marginal utility and probabilistic risk aversion. *Theory and Decision,* **36,** 1–44.

Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann–Morgenstern utilities when probabilities are distorted or unknown. *Management Science,* **42,** 1131–1150.

Wakker, P., & Stiggelbout, A. (1995). Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Medical Decision Making,* **15,** 180–186.

Wakker, P., & Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty,* **7,** 147–175.

Wakker, P. P., Thaler, R. H., & Tversky, A. (1997). Probabilistic insurance. *Journal of Risk and Uncertainty,* **15,** 7–28.

Weber, E. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin,* **115,** 228–242.

Weibull, J. (1982). A dual to the von Neumann–Morgentern theorem. *Journal of Mathematical Psychology,* **26,** 191–203.

Wu, G. (1994). An empirical test of ordinal independence. *Journal of Risk and Uncertainty,* **9,** 39–60.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science,* **42,** 1676–1690.

Wu, G., & Gonzalez, R. (1998). Common consequence conditions in decision making under risk. *Journal of Risk and Uncertainty,* **16,** 115–139.

Wu, G., & Gonzalez, R. (1999). Nonlinear decision weights in choice under uncertainty. *Management Science,* **45,** 74–85.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica,* **55,** 95–115.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika,* **46,** 357–388.

Accepted November 18, 1998