Violations of monotonicity in choices between gambles and certain cash

MICHAEL H. BIRNBAUM
California State University, Fullerton
and Institute for Mathematical Behavioral Sciences

LAURA A. THOMPSON
California State University, Fullerton

Recent studies found that judgments of the values of gambles violated *monotonicity* (a form of dominance). There are gambles for which an increase in the value of one outcome, holding everything else constant, causes a decrease in the judged value of the gamble. This paper replicates and extends recent work with choices between gambles and a fixed set of amounts of money that would be received for certain. A set of relations is defined on choice proportions as follows: $A \succ_c B$ (Gamble A is preferred to B on relation c) if and only if the proportion of choices favoring amount c over Gamble A is less than the proportion choosing c over B. Results replicate previous findings in which these choice-based *certainty equivalents* violate monotonicity. The new relation also reveals systematic violations of monotonicity when c is greater than the minimum outcome of the superior gamble, but not when c is less than this value. This result helps clarify the conditions under which monotonicity is violated.

The principle of dominance can be stated briefly as follows: If two alternatives are otherwise identical, but one gamble has one outcome that is preferred to the corresponding outcome of the other gamble, then the gamble with the better outcome should be preferred. For such pairs of gambles, preference should be a monotonic function of any outcome with the others held fixed, so this principle is also called *outcome monotonicity*. Similarly, if two gambles are identical, except one has a greater probability of a preferred outcome and lower probabilities of less preferred outcomes, the one with the higher probability of the better outcome should be preferred. The term *stochastic dominance* is used to refer to the relation between pairs of gambles such that for any outcome, the probability of a lower outcome, given one gamble, is less than or equal to that given the other gamble. The concept of stochastic dominance combines monotonicity with respect to outcomes and monotonicity with respect to probabilities. This paper involves a pure test of outcome monotonicity.

AMERICAN JOURNAL OF PSYCHOLOGY
Winter 1996, Vol. 109, No. 4, pp. 501–523

Although outcome monotonicity seems a very reasonable axiom for the rational decision maker, recent experiments have found situations in which human judgments appear to violate the principle systematically (Birnbaum, 1992b; Birnbaum, Coffey, Mellers, & Weiss, 1992; Birnbaum & Sutton, 1992; Mellers, Weiss, & Birnbaum, 1992).

Let (x, p, y) represent the binary gamble to receive x with probability p and otherwise receive y. Monotonicity requires that (x_1, p_1, y) is preferred to (x_2, p_1, y) if and only if (x_1, p_2, y) is preferred to (x_2, p_2, y); in other words, if and only if x_1 is preferred to x_2. Birnbaum et al. (1992) found that ($0, .05, $96) receives a higher judgment than ($24, .05, $96), even though $24 is judged better than $0; indeed, ($24, .5, $96) receives a higher judgment than ($0, .5, $96). Similar results were also found with different numerical values, and the same violations were also found when subjects were asked to make judgments from the buyer's, neutral's, or seller's points of view.

Birnbaum et al. (1992) represented judgments of binary gambles, (x, p, y), by the following configural-weight model:

$$U_V(x,p,y) = \frac{a_V S_x(p)\, u(x) + (1-a_V)(1-S_x(\mathrm{p}))\, u(y)}{a_V S_x(p) + (1-a_V)\cdot(1-S_x(p))} \tag{1}$$

where $U_V(x, p, y)$ is the utility of the gamble in point of view, V; a_V is the configural weighting parameter for point of view V; $u(x)$ and $u(y)$ are the utilities of the lower- and higher-valued outcomes $(x < y)$; and $S_x(p)$ is a function of the probability of the lower-valued outcome that depends on value: there are different S functions for $x > 0$ and for $x = 0$. For $.04 < p < .96$, $S_x(p)$ can be approximated by $S_x(p) = .59p + .29$, for $x > 0$; however, for $x = 0$, $S_0(p)$ is approximated by $S_0(p) = .74p + .14$. Note that $S_0(p)$ is less than $S_x(p)$, especially for small values of p. According to this model, monotonicity violations occur because the worst outcome of zero has a much lower weight than a worst outcome that is a small positive amount (for the same low probability).

The term *configural* is used to indicate that the parameter representing a stimulus component may depend on the relationships between that component and others that comprise the stimulus array presented on each trial. Subjective expected utility theory, as interpreted by Edwards (1962), for example, is not configural because the weight of each outcome is independent of the value of the outcome or its relationship to other outcomes in the same gamble. The configural-weight model (Equation 1) allows the weight of an outcome to depend on its rank among the other outcomes in the gamble (Birnbaum, 1974). Therefore, the weight of the same outcome with the same probability can be different in different gambles depending on the other outcomes in those gambles (Birnbaum, 1982; 1992a; 1992b; Birnbaum & Sotoodeh, 1991; Birnbaum & Stegner, 1979; Birnbaum, et al., 1992; Weber, Anderson, & Birnbaum, 1992; Weber, 1994).

Configural-weighting models are closely related to rank-dependent utility theories (Lopes, 1990; Luce, 1992; Luce & Fishburn, 1991; Luce & Narens, 1985; Quiggin, 1982; Tversky & Kahneman, 1992; Yaari, 1987), which were developed independently (see review by Wakker, 1993). Birnbaum's (1974, see p. 559) range model was noted to be a rank-dependent, configural-weight model. Configural weighting allows weights to depend on point of view (Birnbaum & Stegner, 1979) and to differ for neutral- or zero-valued outcomes (T. Anderson & Birnbaum, 1976), which allows configural-weight theory to explain violations of monotonicity.

Among utility theories, configural-weight theory is the least restrictive; rank-dependent utility theory and rank-and sign-dependent utility theories are special cases that allow violations of outcome independence but must satisfy monotonicity. To account for violations of monotonicity, the numerical representation of rank and sign-dependent utility theory (Luce & Fishburn, 1991), for example, would have to be modified to allow different weights for different outcomes, but the foundational assumptions that would imply such a representation have not yet been worked out (R. D. Luce, personal communication, January, 1995). Subjective expected utility theory can be interpreted as a special case of rank-dependent theory in which weights are independent of rank, and so this theory implies independence between outcomes (Birnbaum et. al., 1992). Expected utility theory is a special case of subjective expected utility theory in which subjective probabilities are replaced by objective probabilities; expected value theory is a special case of expected utility theory in which utilities are equal to objective values.

The configural-weight parameters, a_V, explain how the rank order of gambles can change in different points of view. For the seller's point of view, a_V was set to .5, and the values estimated for the neutral's and buyer's points of view were approximately .6 and .7, respectively. Configural-weight theory led to estimated $u(x)$ functions that were invariant with respect to point of view (Birnbaum et al, 1992). This theory also led to estimated $u(x)$ functions that also agree with estimates based on "ratios" and "differences" of riskless utility (Birnbaum & Sutton, 1992).

Equation 1 fit the data of Birnbaum et al. (1992), and it predicted the patterns of monotonicity violations obtained by Birnbaum and Sutton (1992), Birnbaum (1992b), and by Mellers, Weiss, and Birnbaum (1992c). Mellers, et al. (1992c) replicated and extended monotonicity violations using different displays of probability and numerical values; they found that violations of monotonicity in judgment persisted even when real money was used as an incentive. Because Equation 1 normalizes relative weights within each gamble, the relative weight of an outcome can be lower when the same amount of probability is divided among several other outcomes, producing violations of branch independence (Weber, Anderson, & Birnbaum, 1992).

One interpretation of the configural-weighting explanation of monotonicity violations was that subjects adopt a simplifying strategy with two-outcome gambles, so zero outcomes would receive a lower weight only for simple, two-outcome gambles. This interpretation implies that violations of monotonicity should not occur with three-outcome gambles having two nonzero outcomes. Nevertheless, the same equations and approximated parameters successfully predicted violations of monotonicity in a new set of three-outcome gambles (Mellers, Berretty, & Birnbaum, 1995), with the additional assumption that the lowest outcome receives the same absolute weight in both two- and three-outcome gambles, and the other two outcomes each receive the weight that a higher outcome receives in a two-outcome gamble.

Although violations of monotonicity have been found consistently in judgment studies in which the key gambles are judged separately, conditions that facilitate comparisons among the gambles appear to reduce violations (Mellers et al. 1992c). Birnbaum and Sutton (1992) obtained direct choices between gambles as well as judgments of the same gambles; they found that although judgments violated monotonicity when gambles are judged one a time, direct comparisons between gambles that contained a transparent dominance relationship rarely violate monotonicity. Because direct choices yield a different ordering from that obtained from judgment, Birnbaum and Sutton identified their finding as a new type of preference reversal.

Some preference reversals can be reduced when choice (rather than judgment) is used to find *certainty equivalents* (Bostic, Herrnstein, & Luce, 1990). Von Winterfeldt, Chung, Luce, & Cho (in press) found different rates of violations of monotonicity when certainty equivalents were obtained from judgments or from different types of choice procedures. The certainty equivalent is the amount of certain money that is psychologically indifferent to a gamble. When certainty equivalents are obtained by the method of judgment (e.g., "how much would you accept to sell this lottery ticket if you owned it?"), the judged values violate monotonicity. However, using a staircase method in which each gamble receives a different set of comparisons, von Winterfeldt et al. concluded monotonicity violations were less frequent with their procedure than they were in judgment.

Birnbaum (1992b) offered subjects choices between gambles and a list of cash values that was the same for all gambles. By examining how each gamble stacked up against a fixed set of cash amounts, this procedure separates choice from transparent comparison. Using this procedure, Birnbaum (1992b) found that violations of monotonicity persisted even when gambles are ordered by choice-based certainty equivalents (based on comparisons between gambles and sure amounts of money). He also found that the inferred certainty equivalent (the value of cash that it is preferred half the time to the gamble) depends on the distribution of cash values offered for comparison against the gambles, which further adds to the difficulty of comparing certainty equivalents when the set of comparison values for the gambles being compared is different.

The following model is useful for discussing choice between gambles and cash amounts:

$$P(c, G) = F[u(c) - U(G)] \tag{2}$$

where $P(c, G)$ is the probability of choosing the sure cash, c, over the gamble $G = (x, p, y)$; U is a function that assigns an overall utility to each gamble; u is the utility function for money; and F is a monotonic function that maps a given utility difference into a choice probability.

The certainty equivalent, c^*, of gamble G is defined as the value of cash that would be indifferent to the gamble in the sense that it would be preferred half the time; (i.e., the value of c^* for which $P(c^*, G) = 1/2$). Birnbaum (1992b) found that values of c^* violated monotonicity when certainty equivalents are determined by a choice procedure in which each gamble is compared to a fixed set of comparison cash amounts. Birnbaum (1992b) also found that the value of c^* depends on the particular set of comparisons used; higher values of c^* were observed when the average value of the comparisons offered was higher than when the comparisons were lower on the average.

To further explore violations of monotonicity, we will also consider the following set of relations. For each value of c, operationally define the relation, $\succ_c$, as follows:

$$A \succ_c B \text{ if and only if } P(c, A) < P(c, B), \tag{3}$$

where $P(c, A)$ represents the observed choice proportion preferring cash amount c over Gamble A.

If Equation 2 held with a single function F, then all of the relations in Expression 3 should agree (i.e., the comparison between two gambles would be independent of the cash value c). The term *scalability* is defined as the agreement of the relations in Expression 3. However, if F is subscripted for each gamble, then the inferred ordering of gambles in this set of relations can depend on the value of c, violating scalability. We will examine these relations with respect to both monotonicity and scalability in order to gain a better understanding of violations of monotonicity.

EXPERIMENT

METHOD

Overview

The experiment was similar to that of Birnbaum (1992b), except that the comparison amounts were more finely spaced and symmetrically distributed to allow examination of the relations in Expression 3. Instructions read (in part) as follows:

> On each trial, you will be offered a comparison between an amount of money and a gamble, or lottery. Your task is to decide whether you would prefer the money (for sure) or the chance to play the lottery (the gamble).

Each gamble was presented with a fixed set of 21 comparison amounts of certain money. The subject's task was to decide between the gamble and each amount of money. They were to indicate their decisions by circling each value of money that would be preferred to the gamble presented on that trial. This task was repeated for a number of gambles. The gambles were selected to include some that were predicted to show violations of monotonicity on the basis of previous research.

Stimuli and design

Gambles were displayed as in the following example:

.2	.8
\$24	\$96

This display represents a probability of .2 to win \$24 and a probability of .8 to win \$96. Subjects were instructed to imagine a can containing 100 slips of paper, of which 20 slips had "\$24" written on them and 80 specified "\$96"; one slip would be chosen at random to determine the amount won. Probabilities displayed always summed to one.

The 30 binary gambles were generated from a factorial design of six pairs of amounts [(x, y) = (\$0, \$24), (\$0, \$48), (\$0,\$96), (\$24, \$48), (\$24, \$96), (\$48, \$96)], combined with five levels of the probability of receiving the smaller amount [p = .05, .2, .5, .8, or .95]. Each of the 30 gambles was presented for comparison with following amounts: \$1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 99, which were printed in ascending order, in a vertical column below each gamble.

Procedure

The thirty gambles, with their comparison amounts, were printed in random order in booklets; each booklet began with instructions and six warm-up trials.

Instructions stated that subjects should prefer a gamble to any amount less than the least amount the gamble would offer and that they should prefer amounts of money that exceeded the most the gamble could offer. Subjects who violated these properties during the warm-up were directed to reread the instructions before proceeding. (This instruction rules out any response less than $24 for the [$24, *p*, $96] gambles, e.g., but it does not rule out small values for the [$0, *p*, $96] gambles. Therefore, the instruction would tend to enhance the likelihood of satisfying monotonicity).

Subjects

The subjects were 80 undergraduates at California State University, Fullerton, who received extra credit in introductory psychology.

RESULTS

Table 1 shows the mean value of the smallest amount of certain cash (*c*) that was preferred to each of the 30 gambles, as in Birnbaum (1992b). The ordinal pattern of results is similar to that previously reported by Birnbaum (1992b) for two skewed contexts of comparison values. The mean judgments in Table 1 show violations of monotonicity in the same fashion as in the previous research. These violations are illustrated in Figure 1.

Table 1. Mean value of smallest amount of certain cash (*c*) preferred to each gamble

	Values of *p*				
Outcomes	.95	.8	.5	.2	.05
($0, $24)	13.3	15.2	17.8	20.3	21.9
($0, $48)	21.3	25.1	30.9	39.4	43.3
($0, $96)	28.7	29.0	49.2	*71.3*	*77.5*
($24, $48)	31.6	33.1	37.9	40.2	41.8
($24, $96)	39.6	40.7	51.4	*64.1*	*73.4*
($48, $96)	55.7	57.8	64.4	71.0	76.8

Note. Each entry is the mean of the smallest amount of certain cash that is just preferred to each gamble; *p* is the probability to receive the smaller outcome. Italicized entries show violations of monotonicity discussed in text and shown in Figure 1.

Figure 1 plots the mean (of the minimum amount of *c* that was preferred to each gamble) as a function of the probability to win \$96 (1–*p*), with separate curves for (\$0, *p*, \$96) and for (\$24, *p*, \$96). Monotonicity implies that judgments of (\$24, *p*, \$96) should exceed (\$0, *p*, \$96) for all values of *p*; (i. e., the curves should not cross). Instead, the curves cross, and the mean responses of (\$0, *p*, \$96) are actually higher than the means for (\$24, *p*, \$96) when $p = .2$ and $p = .05$ (i.e., $1-p = .8$ and $1-p = .95$). For these two values of *p*, means are significantly *higher* for the gambles with a zero outcome than for the dominant gambles with a lowest outcome of \$24, $F(1, 79) = 8.38$. Of the 80 individual subjects, only 11 subjects gave higher judgments to both of the dominating gambles for these two pairs. In contrast, 47 violated monotonicity at least once for these two pairs, including 19 who violated monotonicity for both of these values of *p* (the rest involved ties, which were not counted as violations). Therefore, the present data replicate the violations of monotonicity observed in previous research.

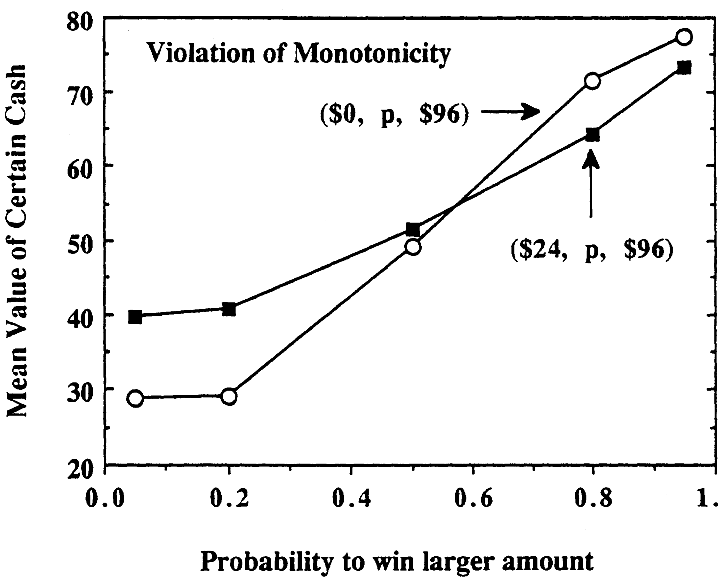

Figure 1. Mean value of the smallest comparison amount of certain cash (*c*) preferred to each gamble, plotted against the probability to receive \$96. Monotonicity implies that the curve for the gamble (\$24, *p*, \$96), shown as solid squares, should always fall above the curve for (\$0, *p*, \$96), shown as open circles; instead, crossing of curves indicates violations of monotonicity

Table 2 presents the percentage of choices of c over the gamble for the 15 gambles in which \$96 was the highest outcome. Each row represents a different gamble, and each column represents a different comparison amount (abbreviated in Table 2 to show c in \$10 increments). The percentages increase from left to right because as one increases the value of certain cash, the more likely it is that the subject will prefer the cash to the gamble. The percentages decrease as the probability to win \$96 increases because as the gamble improves, the tendency to choose the cash is decreased.

Table 2 also contains violations of monotonicity of the relation in Expression 3. For example, examining the rows for A = (\$0, .2, \$96), B = (\$24, .2, \$96), and C = (\$48, .2, \$96), we see that c = \$70 is preferred to (\$0, .2, \$96) only 42% of the time, but \$70 is preferred to (\$24, .2, \$96) and (\$48, .2, \$96) 55% and 54% of the time, respectively. However, for the same value of c, when p = .8, the order of the gambles is monotonic with the value of x. Therefore, by Expression 3,

$$(\$0, .2, \$96) \succ_{70} (\$48, .2, \$96) \succ_{70} (\$24, .2, \$96),$$

but $$(\$48, .8, \$96) \succ_{70} (\$24, .8, \$96) \succ_{70} (\$0, .8, \$96),$$

violating monotonicity for the comparisons of B and C versus A (for c = \$70), but not between the two gambles with positive worst outcomes (\$24 vs. \$48).

Figure 2 plots the percentage of choices favoring the cash over each gamble, plotted as a function of the sure amount, with a separate curve for each level of probability of winning \$48. The data values of P(c, G) appear to increase as a function of c and decrease monotonically as a function of p, for fixed x and y. Similar results were obtained for x = \$96 with y fixed, for varying values of p, as can be seen in Table 2. However, when different values of x are plotted in the same panel, as in Figure 3, the violations of monotonicity become apparent.

Figure 3 explores the violations of monotonicity by plotting the proportion of choices favoring the money (as in Figure 2) for A = (\$0, .2, \$96) and B = (\$24, .2, \$96). Note that Gamble B dominates A, since the outcomes of B are the same or better than those of A. The certainty equivalent is defined as the projection onto the abscissa of the intersection of the data curves and the ordinate value of 50%. As shown in the figure (see vertical arrows), certainty equivalents for these gambles violate monotonicity: A has a higher certainty equivalent than B. Similar results were obtained for p = .05 (see Table 2). For values of c greater than or equal to \$25, P(c, B) > P(c, A), indicating that for these values, $A \succ_c B$, in violation of monotonicity. Note that for values of c below \$24,

Table 2. Percentage of choices of certain cash (c) over gamble

	Value of c								
Gamble	$10	$20	$30	$40	$50	$60	$70	$80	$90
($0, .95, $96)	42	64	74	76	81	82	85	85	88
($24, .95, $96)	0	1	59	75	84	85	88	90	91
($48, .95, $96)	0	0	0	2	64	88	90	91	94
($0, .80, $96)	32	52	66	79	86	88	91	91	94
($24, .80, $96)	0	4	51	70	84	88	88	90	91
($48, .80, $96)	0	0	0	2	58	81	86	89	92
($0, .50, $96)	11	18	25	46	66	76	78	80	84
($24, .50, $96)	0	2	29	38	66	78	81	82	86
($48, .50, $96)	0	0	0	1	35	58	75	81	89
($0, .2, $96)	0	5	10	19	29	34	42	58	69
($24, .2, $96)	0	1	19	28	42	46	55	62	76
($48, .2, $96)	0	0	0	1	26	44	54	68	78
($0, .05, $96)	0	1	6	10	24	28	32	41	59
($24, .05, $96)	0	0	12	19	26	34	41	48	62
($48,. 05, $96)	0	0	0	1	26	35	39	52	60

Note. Each entry is the percentage of choices preferring c over the gamble; that is, data correspond to 100 times P(c, G) of Equation 2. Each row designates a different gamble (G), and each column indicates a different value of c. Table has been simplified by including only nine cash amounts in $10 increments (there were actually 21 values).

the relation reverses. These reversals of preference due to changes in c, called *violations of scalability*, are further explored in Figure 4.

Figure 4 shows corresponding results for ($0, .2, $96) and ($48, .2, $96), which shows the violations of scalability more clearly. Figure 4 shows that for values of c less than $48 (the worst outcome), monotonicity of $\succ_c$ is satisfied, but when $c > \$55$, monotonicity is systematically violated. For these gambles, the means (in Table 1) are very similar for the two gambles; however, the certainty equivalents (values of c^* that would correspond to abscissa projections of 50%) and the $\succ_c$ relationship (for values of $c > \$55$) show a marked violation of monotonicity in Figure 4.

Figure 4 appears consistent with the idea that the variability of judgment that produces the slopes of the curves in Figures 2, 3, and 4 is produced by variation in the weights of the two monetary amounts. The greater the range of outcomes, the lower the slope. Similar results were also found for $p = .05$ (see Table 2) and for ($24, p, $48) for the same levels of p.

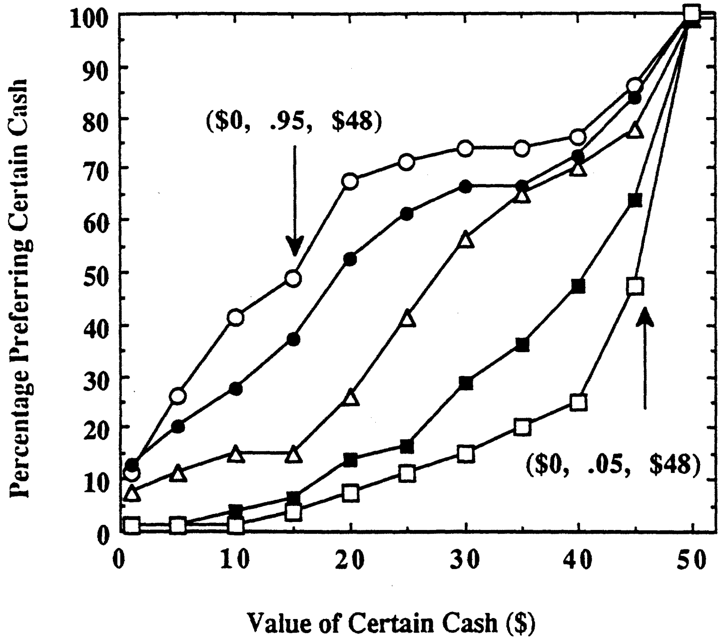

Figure 2. Percentage of choices favoring the certain cash (c) over the gamble (with a probability of p to win $0, otherwise $48), plotted as a function of c. Separate curves are shown for each value of p. Inferred certainty equivalents are projections on the abscissa corresponding to the ordinate value of 50%

DISCUSSION

These results add to a growing literature in judgment and decision making of puzzling phenomena that trouble the theoretician. It will be useful for the sake of discussion to review briefly some of the premises and procedures of judgment and decision making, to see how recent results have created difficulty for early theories and to show how the present results fit into the picture.

Axioms and "rationality"

One source of consternation in judgment and decision making has arisen because the term *axiom* has two meanings in these discussions (see e.g., Tversky & Kahneman, 1986; Luce, 1986). Initially, the idea of an axiom was that of an unquestioned, self-evident truth from which we deduce other truths. It would be rational to satisfy the consequences of the axioms, if indeed the axioms are self-evident and the deductions

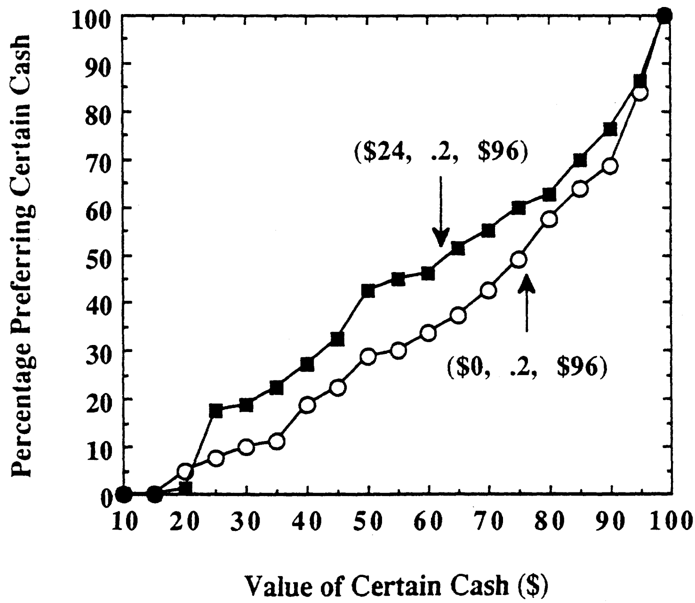

Figure 3. Percentage of choices favoring certain cash (*c*) as a function of *c*, with separate curves for two gambles: ($0, .2, $96), shown as open circles, and ($24, .2, $96), shown as solid squares. Monotonicity implies that solid squares should be below the open circles; instead, for values of *c* above $25, the dominated (worse) gamble appears to be more likely preferred against the cash. Similar results were obtained for $p = .05$

logical. However, when educated and seemingly rational people found that their own decisions violated the axioms, arguments concerning the rationality of certain axioms developed. Some scholars were troubled by the discrepancy between the deductions from the axioms and what they regarded as their own reasoned decisions (e.g., Savage, 1954). Other scholars decided that they would prefer to satisfy their own preferences and let the axioms satisfy themselves.

In later developments, the term axiom came to refer to theoretical premises from which predictions of behavior could be deduced, with the original ("self-evident truth") meaning of the term axiom converted to a purely descriptive, first premise. If humans' decisions were consistent with the descriptive axioms, then behavior can be explained, but the new descriptive premises need not be considered completely rational by the first definition. Such an explanation would be analogous to the explanations of the Ames room illusion and size constancy using

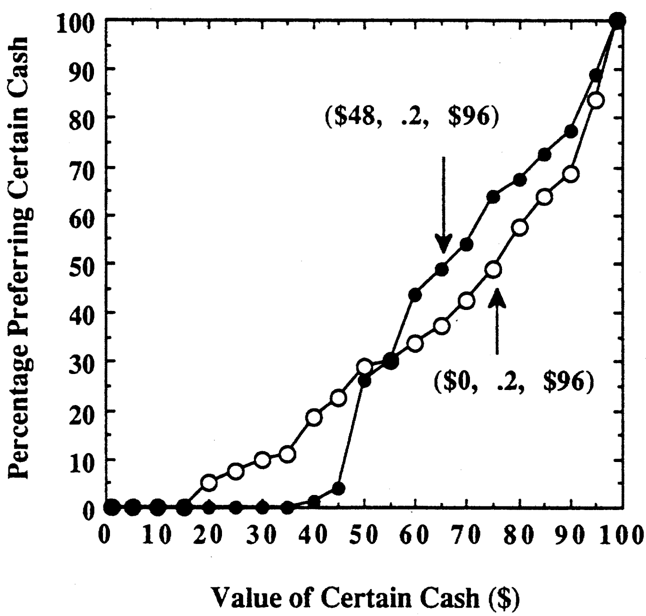

Figure 4. Percentage of choices preferring the certain cash (c) to two gambles: ($48, .2, $96) shown as solid circles, and ($0, .2, $96) shown as open circles, as in Figure 3. When the solid circles are above the open circles, there is a violation of monotonicity for that value of c; crossing of the curves indicates a violation of scalability. In this case, monotonicity is satisfied for values of $c <$ \$55, but not for values of $c >$ \$55. The certainty equivalents (abscissa projections for 50% ordinate) also violate monotonicity. Similar results were obtained for $p =$ .05, and for (x, y) = ($24, $48) versus ($0, $48)

the same size constancy mechanism (Birnbaum, 1983; Tversky & Kahneman, 1986). The same psychological theory might explain behavior that is deemed "correct" (constancy) and deemed "erroneous" (illusions).

Let us start at the beginning. As psychologists, we assume that different outcomes can be compared with respect to their desirability. Most people would prefer a million dollars to a movie ticket, and they would prefer the ticket to a hit in the head; so we hypothesize that things such as amounts of money, entertainments, and hedonic experiences can be compared and scaled according to a psychological scale that we will call *utility*. Different people might have different utilities for the same things, so that two people might willingly exchange their possessions and both feel better off after the trade.

The standard theory of economics begins with the story of two people on a desert island: Sam with ten cans of beer and Janet with ten bags of potato chips. Each one wants to trade, and they freely choose to exchange beers for potato chips until Janet has six cans of beer and four bags of potato chips, and Sam has four cans of beer and six bags of potato chips. At this point, no further trading occurs. Presumably, both people have improved their situations by trading; when they cannot agree on a further trade, it is presumed that no trade would increase both utilities. This classic problem in economics was investigated empirically by Thurstone (1931) and has continued to interest psychologists (Shanteau & Troutman, 1992). If the utility of a number of goods is a negatively accelerated function (the first beer has more utility than the second, which has more than the third), people will tend to exchange goods they possess in quantity for goods that they lack.

The concept of utility, in particular, that utility of money might be a negatively accelerated function of actual money (as proposed by Bernoulli and later Fechner), was also useful in explaining why people seem to be risk averse, or prefer the sure-thing expected value of a gamble to the gamble itself. This concept of utility was formalized by von Neumann & Morgenstern (1947) and later expanded by Savage (1954) to include two psychological concepts, utility and *subjective probability*. The formal theories involved the idea of a *preference relation,* which if it satisfied certain axioms, could be used to measure the two psychological constructs (Krantz, Luce, Suppes, & Tversky, 1971).

As psychologists, we try to learn something about these psychological values by asking subjects to rate, evaluate, judge, or compare alternatives. Unfortunately, different techniques that seem to ask the same question yield different answers. For example, we can ask subjects to rate how attractive each gamble is on a rating scale of attractiveness. We could ask each subject to judge the most they would pay to buy each gamble. We can ask subjects to judge the least they would accept to sell each gamble. We can ask subjects to compare gambles in pairs and indicate which they would prefer within each pair. We can ask subjects how much they would pay to get one gamble rather than another.

These different operational definitions of the preference relation lead to different preference orders (Lichtenstein & Slovic, 1971; Lindman, 1971; Mellers, Ordóñez, & Birnbaum, 1992b; Mellers, Chang, Birnbaum, & Ordóñez, 1992a). Such preference reversals were both exciting and upsetting to theorists because the most fundamental idea of theories of psychology and economics is the preference order (von Winterfeldt & Edwards, 1986; Luce, 1992; Slovic, Lichtenstein, & Fischhoff, 1988; Stevenson, Busemeyer, & Naylor, 1991). These theories start with the idea that people have values for such things as four cans of beer, a million dollars, or a hit in the head; the idea that the preference relation itself is malleable or unstable is troublesome.

We can state five axioms of preference: Consistency, Transitivity, Scalability, Monotonicity and Outcome Independence. These are all similar in that they hope to define a consistent account of preferences but have been violated in different recent experiments.

1. *Consistency* is defined as the property that the preference order should be independent of the method used to elicit it. Preference reversals between different methods violate consistency. For example, it is possible to find Gambles, A and B, such that people offer to pay more to play Gamble A than B, but in paired comparison, people choose to play B rather than A (Lichtenstein and Slovic, 1971; Lindman, 1971). A number of theories have been proposed to account for preference reversals between different methods; these theories are contrasted in Mellers et al. (1992a; 1992b). The premises of the theories offered to explain the results are psychological theories that are intended to be empirically descriptive.

2. *Transitivity* is defined as follows: If A is preferred to B, and B is preferred to C, then A is preferred to C. It seems hard to debate transitivity as a rational principle. If a person systematically and repeatedly violated transitivity, that person could apparently be made into a "money pump." Presumably, they would pay money to exchange B for C, pay to get A instead of B, pay more to get C instead of A, pay again to get B instead of C, and so on forever. This conclusion depends on a particular definition of preference that supposes that people would pay money to go from one state to a more preferred state. We see below that there is a subtle interplay between formal ideas, operational definitions, and empirical results.

For example, if we ask people to assign a price to each gamble, the data automatically satisfy transitivity because the subjects assign real numbers to the gambles, and numbers obey transitivity. However, if people are presented with gambles in pairs, and are asked to choose between A and B, choose between B and C, and choose between A and C, the choices may or may not satisfy transitivity. To distinguish systematic violations of transitivity from momentary fluctuations of judgment, (weak stochastic) transitivity has been defined for choice proportions as follows:

$$\text{If } P(A, B) > \tfrac{1}{2}\text{ , } P(B, C) > \tfrac{1}{2}\text{, then } P(A, C) > \tfrac{1}{2}$$

where $P(A, B)$ is the probability of choosing A over B in a paired comparison.

3. *Scalability,* a stronger form of transitivity, (also called *strong stochastic transitivity*) is defined as follows:

If $P(A, B) > \frac{1}{2}$, $P(B, C) > \frac{1}{2}$, then $P(A, C) >$ maximum of $[P(A, B), P(B, C)]$

For the present data, the analogous property (that is violated in the present data) can be stated as follows:

$$P(A, c_1) > P(B, c_1) \text{ if and only if } P(A, c_2) > P(B, c_2).$$

The violation can be seen in the crossing of the curves in Figure 4. Violations of strong stochastic transitivity have been observed when the stimuli to be compared differ in their similarity (Tversky, 1969; Mellers & Biagini, 1994).

To explain the choice proportions obtained in this experiment, Equation 2 can be retained only if the function, F, is permitted to depend on the range of outcomes in the gamble. The crossovers of the curves in Figure 4 constitute violations of scalability, the premise that all of the relations (of Expression 3) should have agreed. Busemeyer (1985) also found violations of scalability that constitute evidence against Equation 2 unless the function, F, were permitted to depend on the variance of the outcomes in each gamble. Thus, the choice relations, $\succ_c$, do not define a scalable order of gambles for all values of c. Instead, the inferred order of gambles depends on the relationship between the value of c and the outcomes of the gamble, as shown in Table 2 and illustrated clearly in Figure 4.

4. *Monotonicity* assumes that the ordering due to one outcome is independent of other features of the gamble. For this paper, the property of monotonicity that is violated can be stated as follows: (x_1, p_1, y) is preferred to (x_2, p_1, y) if and only if (x_1, p_2, y) is preferred to (x_2, p_2, y). This property seems reasonable as a self-evident axiom, since it seems rational that if we prefer \$24 to \$0 when $p = .8$, then we should make the same preference when $p = .2$. However, the crossing of the curves in Figure 1 reveals a violation. It seems hard to construct a rationale for violating this type of monotonicity. Indeed, when subjects are shown their violations of monotonicity, they do not try to defend them in the same way as they defend discrepancies between buyer's and seller's prices or even the different preference orders obtained with choice and judgment tasks. Instead, they explain them as "errors" of judgment that would not occur in a direct choice (Mellers, et al., 1995).

The present data replicate previously observed violations of monotonicity with new subjects and a symmetric distribution of comparison amounts. These data also show that monotonicity is violated for some, but not all, of a set of relations defined on the probability of choosing each amount of money over each gamble. The data appear consistent with instructions to obey one type of monotonicity: It is rare to prefer a certain cash amount less than the smallest outcome of the gamble. Nevertheless, the new results also reveal violations of monotonicity for cash values greater than the least amount, as in Figure 4. Thus, choice proportions violate both monotonicity and scalability.

The certainty equivalents are consistent with the pattern predicted by configural-weight theory (Equation 1), which summarizes previous results. Configural-weight theory predicts violations from the assumption that the average configural weight of a zero outcome is lower than that of a nonzero, positive outcome of the same probability. It seems reasonable to suppose that different subjects (or the same subject on different occasions) have different values of the configural weights, generating the slopes of curves as in Figures 2–4: the greater the range of outcomes in the gamble, the lower the slope.

Combining the present results with those of Birnbaum & Sutton (1992), the present pattern of monotonicity violations seems to imply a violation of transitivity as well. Birnbaum and Sutton found that when given a transparent choice between B = (\$24, .2, \$96) and A = (\$0, .2, \$96), people prefer the dominating gamble B to A, so $P(B, A) > .5$. However, Table 2 shows that if C = \$70, then we have the following intransitive conclusion: $P(B, A) > .5$, $P(A, C) = .58$, but $P(C, B) = .55$. Thus, B is preferred to A, A to C, but C is preferred to B. It is unclear if such cross-experiment comparisons would predict what a single individual would do when faced with all three comparisons, but it should be clear that the theoretician has a problem to account for all of the data in terms of a single, transitive preference order.

(5) *Outcome independence* is similar to monotonicity in that it seeks to define a consistent preference ordering; however, this principle's a priori rationality is still a subject of debate (see Savage, 1954; Birnbaum, et al., 1992). For gambles consisting of three, equally likely outcomes, outcome independence can be written as follows:

$$(x, y, z) \text{ is preferred to } (x', y', z)$$
$$\text{if and only if}$$
$$(x, y, z') \text{ is preferred to } (x', y', z')$$

Basically if the combination of (x,y) is preferred to (x',y') when the common consequence is z, then the same preference order should hold when the common consequence is z'. Rank-dependent utility theories, including configural-weight theory, do not require this property to be satisfied, although subjective expected utility theory (and any special case) does require it (Savage, 1954). Birnbaum, et al. (1992) and Weber et al. (1992) found violations of this property (also called *branch independence*) in judgment experiments. An experiment by Wakker, Erev, & Weber (1994) distinguished two types of branch independence, which they called *comonotonic independence* (when z and z' have the same rank in both choices) and *noncomonotonic independence* (when the rank of z and z' are different). Their experiment found that both types of branch independence were fairly well satisfied in a choice experiment. Birnbaum and McIntosh (1996) found violations of branch independence in choices that violated the predictions of cumulative prospect theory, according to the weighting function estimated by Tversky and Kahneman (1992); Birnbaum and Beeghley (in press) found a similar pattern of violations using judgments of buying and selling price. Preferences, buying prices, and selling prices could be fit by configural-weight theory using the same utility function for money, allowing the weight of each outcome to depend on its rank and the task.

Judgment versus choice

Formally, it is difficult to distinguish judgment and choice. In a judgment experiment, stimuli are usually presented one at a time, and the subject assigns a judgment (assigns a value) to each. Presumably, the subject evaluates each stimulus separately, so it seems that it should not matter what stimuli are included for judgment. However, in a judgment task, the subject must choose the "best" response from a set of responses (that might be infinite), so judgment can be interpreted as a decision task in which the subject makes a judgment by choosing among potential responses. Furthermore, it turns out that the judgment of a given stimulus does reflect comparisons with other stimuli because the judgment of any stimulus depends on the set of other stimuli presented with it. For example, Mellers, et al. (1992b) found that the judged value of a gamble might be \$10 higher when it was presented among gambles of mostly lower expected values than it was when presented among gambles of mostly higher expected values. Thus, so-called "absolute" judgments are relative, in that the judgments of any given stimulus contain (implicit) comparisons of that stimulus against the others that were presented for judgment, and each response involves (implicit) comparisons of that response against the set of available responses.

Similarly, in a choice experiment, it is often presumed that each comparison could be made independently of the other choices offered, so choices seemingly "avoid" the contextual effects observed in judgment studies. However, theoretically, one can represent choice proportions as the mean judgments of a two-category scale of preference between two stimuli and analyze choice proportions in the same fashion as any other judgment. Empirically, it turns out that choice proportions between stimuli depend on the distribution of choices offered. For example, Birnbaum (1992b) found that same amount of money was either preferred to a given gamble or not, depending on whether other amounts offered for comparison were mostly larger or mostly smaller than the given amount.

Intuitively, procedures of judgment and choice seem to fall on a continuum. At one extreme, subjects can be asked to judge the certainty equivalents of gambles from various points of view, as in Birnbaum et al. (1992) and Mellers et al. (1992c). In such experiments, subjects do not necessarily directly compare gambles that may possess a dominance relation, since they assign the values to the gambles on different trials and presumably do not remember the exact values they have assigned. A number of experiments with such procedures find that monotonicity can be systematically violated in such experiments.

At the other extreme on this intuitive continuum are experiments that offer direct choices between gambles with a "transparent" dominance relationship (Tversky & Kahneman, 1986). Birnbaum and Sutton (1992) found that violations of dominance in this situation are rare, even for the same subjects who violated monotonicity in judgments of the same stimuli. Indeed, Mellers et al. (1992c) and Mellers et al. (1995) asked subjects to comment on their prices for two such gambles and found that subjects considered their own judgments to be in error when they found (after the experiment) that they had violated dominance.

The choice-based certainty equivalent method seems intuitively to be an intermediate method between direct choice and direct judgment. The procedure formally involves choice, but the gambles are contrasted to money and are only compared indirectly with each other. Because each gamble is given a numerical value, the results must satisfy transitivity, as is the case in a judgment experiment.

Indeed, Birnbaum's (1992b) variation of this choice procedure (also used in the present study) seems intuitively to be even closer to judgment because of the simultaneous presentation of the list of comparison values. On one hand, the procedure formally asks for a series of choices between fixed amounts and gambles, so it is clearly a choice experiment. On the other hand, a subject may choose a judgment of the worth of the gamble and then fill in all of the choices. Additionally, isn't it reasonable to suppose that *any* judgment is the consequence of a series of implicit choices made by the subjects? Indeed, instructions for a judgment task often include explicit choices to illustrate how to make a judgment ("How much would you pay for this gamble? Would you pay $10? Would you pay $20? Well, what is the most you would pay?"). Yet despite these seeming similarities and presumed formal equivalence, the results of the different procedures can be different.

Another procedure to determine choice-based certainty equivalents would present each comparison on a different trial. Staircase procedures use such a method, which separates the comparisons by intervening trials but also confounds the stimulus with the set of comparisons. The choice of a confound is based on implicit (but untested) theories of how to use each subject's response to each comparison to choose the next value to present. The basic idea is similar to the idea used in computerized adaptive testing. The untested assumption is that either there is an underlying function, F, that is either independent of the context of comparison stimuli, and each comparison represents evidence of the same F function, or that the set of comparisons produced by the algorithm is the "right" set.

Von Winterfeldt et al. (in press) used several procedures to assess certainty equivalents for gambles used by Mellers et al. (1992c). They reported fewer violations of monotonicity when certainty equivalents were determined by a variant of the Parameter Estimation by Sequential Testing (PEST) procedure than with other procedures. In contrast with the present procedure, in PEST, each gamble is presented with a single comparison amount on each trial. Different gambles are intermixed, and the selection of the next comparison value depends on the subject's previous response, according to a staircase algorithm designed to hone in on the certainty equivalent for each gamble. This procedure differs from the present procedure in that the set of comparison values is thus confounded with the gamble and the subject, whereas in the present procedure, the set of comparisons was the same for all of the gambles compared. Unfortunately, the PEST algorithm causes a gamble of higher expected value to receive comparisons of higher average value than a gamble of lower expected value. Such a procedure may thus find greater satisfaction of monotonicity because it capitalizes on contextual effects and the monotonicity of expected values, rather than because the procedure itself reveals a "truer" measure of certainty equivalents.

Procedures that are formally judgment tasks but facilitate comparisons among the gambles (e.g., the short version of Mellers et al., 1992c, which used a small number of stimuli presented for judgment on the same page) apparently lead to greater consistency with monotonicity. Thus, we think that the key to reducing violations of monotonicity lies in making the dominance relation between the gambles clear, rather than using choice or judgment as the task.

Notes

This research was supported by a National Science Foundation Grant, SES 8921880, to the senior author through California State University, Fullerton Foundation and facilitated by SBR-9410572. This project was also supported by a Summer Research Grant from the State of California. We thank R. D. Luce, D. von Winterfeldt, and Younghee Cho for useful discussions of these topics and R. D. Luce for helpful comments on an earlier draft. Received for publication August 22, 1994; revision received January 3, 1996.

Correspondence concerning this article should be addressed to Michael H. Birnbaum, Department of Psychology, California State University, P.O. Box 6846, Fullerton, CA 92834-6846 (E-mail: mbirnbaum@fullerton.edu). Laura A. Thompson is now a graduate student at Baylor University, Waco, TX.

References

Anderson, T., & Birnbaum, M. H. (1976). Test of an additive model of social inference. *Journal of Personality and Social Psychology, 33,* 655–662.

Birnbaum, M. H. (1974). The nonadditivity of personality impressions [Monograph]. *Journal of Experimental Psychology, 102,* 543–561.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.

Birnbaum, M. H. (1983). Scale convergence as a principle for the study of perception. In H.G. Geissler & V. Sarris (Eds.), *Modern issues in perceptual psychology* (pp. 319–335). Amsterdam: North Holland.

Birnbaum, M. H. (1992a). Issues in utility measurement. *Organizational Behavior and Human Decision Processes, 52,* 319–330.

Birnbaum, M. H. (1992b). Violations of monotonicity and contextual effects in choice-based certainty equivalents. *Psychological Science, 3,* 310–314.

Birnbaum, M. H., & Beeghley, D. (in press). Violations of branch independence in judgments of the value of gambles. *Psychological Science.*

Birnbaum, M. H., Coffey, G., Mellers, B. A., & Weiss, R. (1992). Utility measurement: Configural-weight theory and the judge's point of view. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 331–346.

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

Birnbaum, M. H., & Sotoodeh, Y. (1991). Measurement of stress: Scaling the magnitudes of life changes, *Psychological Science, 2,* 236–243.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37,* 48–74.

Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes, 52,* 183–215.

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

Busemeyer, J. R. (1985). Decision making under uncertainty: Simple scalability, fixed sample, and sequential sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 583–564.

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

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

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

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

Lopes, L. L. (1990). Re-modeling risk aversion: A comparison of Bernoullian and rank dependent value approaches. In G. M. von Fuerstenberg (Ed.), *Acting under uncertainty: Multidisciplinary conceptions.* Boston, MA: Kluwer.

Luce, R. D. (1986). Comments on Plott and on Kahneman, Knetsch, and Thaler. *Journal of Business, 59,* S337–S343.

Luce, R. D. (1992). Where does subjective expected utility fail descriptively? *Journal of Risk and Uncertainty, 5,* 5–27.

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

Mellers, B. A., Berretty, P. M., & Birnbaum, M. H. (1995). Dominance violations in judged prices of two- and three-outcome gambles. *Journal of Behavioral Decision Making, 8,* 201–216.

Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review, 101,* 505–518.

Mellers, B. A., Chang, S., Birnbaum, M. H., & Ordóñez, L. (1992a). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 347–361.

Mellers, B. A., Ordóñez, L., & Birnbaum, M. H. (1992b). A change of process theory for contextual effects and preference reversals in risky decision making. *Organizational Behavior and Human Decision Processes, 52,* 331–369.

Mellers, B. A., Weiss, R., & Birnbaum, M. H. (1992c). Violations of dominance in pricing judgments. *Journal of Risk and Uncertainty,* 5, 73–90.

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

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Shanteau, J. , & Troutman, C. M. (1992). A psychophysical evaluation of diminishing returns in riskless decision making. *Organizational Behavior and Human Decision Processes, 52,* 569–579.

Slovic, P., Lichtenstein, S., & Fischhoff, B. (1988). Decision making. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (Vol. 2, pp. 673–738). New York: Wiley.

Stevenson, M. K., Busemeyer, J. R., & Naylor, J. C. (1991). Judgment and decision-making theory. In M. Dunnette & L. M. Hough (Eds.), *New handbook of industrial-organizational psychology* (Vol. 1, pp. 283–374). Palo Alto, CA: Consulting Psychologist Press.

Thurstone, L. L. (1931). The indifference function. *Journal of Social Psychology, 2,* 139–167.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76,* 31–48.

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 representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297–324.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2d ed.). Princeton: Princeton University Press.

von Winterfeldt, D., Chung, N., Luce, R. D., & Cho, Y. (in press). Tests of consequence monotonicity in decision making under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

von Winterfeldt, D. & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge: Cambridge University Press.

Wakker, P. (1993). Additive representations on rank-ordered sets, II. The topological approach. *Journal of Mathematical Economics, 22,* 1–26.

Wakker, P., Erev, I., & Weber, E. U. (1994). Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty, 9,* 195–230.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effects of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin, 114,* 228–242.

Weber, E. U., Anderson, C. J., & Birnbaum, M. H. (1992). A theory of perceived risk and attractiveness. *Organizational Behavior and Human Decision Processes, 52,* 492–523.

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

References

Comments on Plott and on Kahneman, Knetsch, and Thaler
R. Duncan Luce
The Journal of Business, Vol. 59, No. 4, Part 2: The Behavioral Foundations of Economic Theory. (Oct., 1986), pp. S337-S343.
Stable URL:
http://links.jstor.org/sici?sici=0021-9398%28198610%2959%3A4%3CS337%3ACOPAOK%3E2.0.CO%3B2-Z

Rational Choice and the Framing of Decisions
Amos Tversky; Daniel Kahneman
The Journal of Business, Vol. 59, No. 4, Part 2: The Behavioral Foundations of Economic Theory. (Oct., 1986), pp. S251-S278.
Stable URL:
http://links.jstor.org/sici?sici=0021-9398%28198610%2959%3A4%3CS251%3ARCATFO%3E2.0.CO%3B2-C

The Dual Theory of Choice under Risk
Menahem E. Yaari
Econometrica, Vol. 55, No. 1. (Jan., 1987), pp. 95-115.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198701%2955%3A1%3C95%3ATDTOCU%3E2.0.CO%3B2-I