UTILITY INVARIANCE DESPITE LABILE PREFERENCES

Barbara A. Mellers, Elke U. Weber, Lisa D. Ordóñez, and Alan D. J. Cooke

I. Introduction

For over two decades, researchers in decision theory have been puzzled by inconsistencies in human judgment and choice known as *preference reversals.* Preference reversals were first demonstrated in studies of risky decision making by Lichtenstein and Slovic (1971) and Lindman (1971). In these studies, subjects were presented with pairs of gambles matched on expected value. One gamble, referred to as the P-Bet, had a large probability of winning a relatively small amount, and the other gamble, referred to as the $-Bet, had a small probability of winning a relatively large amount. Subjects selected the gamble they preferred from each pair. Then, they assigned selling prices to each separate gamble. The surprising result was that their preferences for gambles derived from choices differed from those derived from selling prices. Subjects chose the P-Bet over the $-Bet, but they assigned a higher selling price to the $-Bet than the P-Bet. Furthermore, the opposite reversal rarely occurred, so the rank order changes could not be attributed to noise or random error.

What do these reversals imply about the measurement of preference? When the weights of objects are measured, one does not expect to find two different orderings depending on whether a balance scale or a digital scale is used. If the digital scale gave systematically higher readings, one would assume the scales were poorly calibrated. But if one object weighed more than another on the balance scale, and the opposite order was obtained with the digital scale, one would worry about the usefulness of the scales. Similarly, when preferences reverse depending on the response mode, it leads one to wonder about the methods of elicitation and the construct itself.

We begin this chapter by examining the robustness of preference reversals. Then, we discuss research in the domain of risky decision making that goes beyond simple demonstrations of preference reversals based on a few pairs of gambles. We show how the entire preference order over a large set of gambles changes with the response mode. In addition, we examine preference reversals in a riskless domain in which subjects state their preferences for apartments using choices and ratings of attractiveness. We propose a two-pronged theory of preference reversals. In the risky domain, we attribute preference reversals to different decision strategies across tasks. In the riskless domain, we argue that preference reversals are caused by weight changes across tasks, a variation of a theory proposed by Tversky, Sattath, and Slovic (1988): The attribute judged more important has a greater effect in choices than in ratings. Finally, we show how this two-pronged theory, which assumes that subjects change either strategies or weights across tasks, gives a coherent account of many important properties of the data while allowing the elicitation of utilities or psychological values to remain constant across tasks. We demonstrate that, with the appropriate models, utilities are stable and have meaning over and beyond the task from which they are derived.

II. Robustness of Preference Reversals

The first experiments on preference reversals gave rise to a large number of studies that tested the robustness of the phenomenon in both risky and riskless domains. Although preference reversals in risky domains were demonstrated primarily with gambles, other stimuli were also used. Mowen and Gentry (1980) presented subjects with products having either high probabilities of success or high expected profits. Subjects chose products with high probabilities of success, but assigned higher selling prices to products with higher expected profits. Reversals of preference were also found in financial settings. Tversky, Slovic, and Kahneman (1990) showed that people chose the short-term investment with a lower yield over the long-term investment with the higher yield, but they assigned a higher price to the long-term investment.

Preference reversals were also demonstrated in the riskless domain. Bazerman, Loewenstein, and White (1992) showed that people reversed their preferences for reward allocations using choices and ratings. Subjects chose the allocation of $600 for oneself and $800 for the other person over the allocation of $500 for both parties, but they rated the latter as more desirable than the former. Irwin, Slovic, Lichtenstein, and McClelland (1993) demonstrated the reversals with public policy questions. They asked people about their willingness to pay for improvements in air quality and improvements in consumer goods (e.g., a better camera or VCR). Irwin et al. found that, in a choice task, people paid more for improvements in air quality than for improvements in consumer goods; but in a pricing task, people assigned a higher value to improvements in consumer goods. Finally, Chapman and Johnson (1993) obtained preference reversals with health items and commodities. Health items were judged as more valuable when measured with life expectancy evaluations, but commodities were judged more valuable with monetary evaluations.

Not only have preference reversals been found in risky and riskless domains, they also occur with different response modes. In addition to selling prices and choices, preference reversals were demonstrated with choices and attractiveness ratings (Goldstein & Einhorn, 1987), attractiveness ratings and buying prices (Goldstein & Einhorn, 1987), and buying prices and selling prices (Birnbaum & Sutton, 1992), among other combinations.

Numerous attempts have been made to reduce preference reversals or eliminate them entirely. Financial incentives have had either limited effects (Pommerehne, Schneider, & Zweifel, 1982) or no effect at all (Lichtenstein & Slovic, 1971; Grether & Plott, 1979). Even when subjects used their *own* money to gamble in a casino in Las Vegas, preference reversals continued at approximately the same rate as found in laboratory studies (Lichtenstein & Slovic, 1973). Extensive instructions have also been used to reduce preference reversals with mixed success. Lichtenstein and Slovic (1971) gave "lengthy and careful" instructions to their subjects, but the phenomenon continued to occur. Reilly (1982), however, taught subjects about expected values and significantly reduced the rate of preference reversals. Other researchers have examined the effects of multiple plays on preference reversals. Wedell and Böckenholt (1990) told people to assume that they were playing each gamble 100 times and found that preference reversals with choices and selling prices were significantly reduced, but not eliminated.

Experimental markets, where buyers and sellers make trades using specified rules of communication, have had more success at reducing preference reversals (Cox & Grether, 1992). A particularly powerful technique is a money pump. With a money pump, people are caught in a cycle of exchanges in which they eventually end up with their initial option *and* less money.[1] Berg, Dickhaut, and O'Brien (1985) found that subjects still made preference reversals after one cycle of a money pump. However, Chu and Chu (1990) showed that preference reversals were completely extinguished after three cycles. In sum, certain types of feedback from the market may correct the inconsistencies, but very few decisions have such immediate and obvious feedback.

III. Comparing Preference Orders

We conducted several experiments designed to compare preference orders across a variety of different tasks. In Mellers, Chang, Birnbaum, and Ordóñez (1992) and Mellers, Ordóñez, and Birnbaum (1992), we used 36 two-outcome gambles of which one outcome was a positive payoff and the other outcome was zero. The probabilities, payoffs, and expected values for these gambles are presented in Fig. 1. Probabilities and payoffs were selected so that gambles along the diagonals would have the same expected values.

In this section we focus on responses from three tasks: selling prices (minimum amounts that people would accept to sell the gambles), attractiveness ratings, and strength of preference judgments (choices, followed

PROBABILITY \ PAYOFF	$3.00	$5.40	$9.70	$17.50	$31.50	$56.70
0.05	0.15	0.27	0.49	0.88	1.58	2.84
0.09	0.27	0.49	0.87	1.58	2.84	5.10
0.17	0.48	0.86	1.55	2.80	5.04	9.07
0.29	0.87	1.57	2.81	5.08	9.14	16.44
0.52	1.56	2.81	5.04	9.10	16.38	29.48
0.94	2.82	5.08	9.12	16.45	29.61	53.30

Fig. 1. Expected values for two outcome gambles with some probability (rows) of winning a specified amount (columns), or otherwise zero.

---

[1] Given a choice between a P-Bet and a $-Bet, subjects often choose the P-Bet. Subjects then pay to exchange the P-Bet for the $-Bet (as the buying price for the $-Bet is higher than the P-Bet). When given the choice, subjects then trade the $-Bet for the P-Bet. The result is that subjects have incurred a loss to regain the initially owned gamble.

by judgments of the magnitude of preference). In each task, we derived preference orders, which are shown in Fig. 2. The numbers 1 and 36 represent the least preferred and most preferred gambles, respectively. Arrows indicate the direction of preference for pairs of gambles with equal expected values. Arrows pointing upward show pairs for which the P-Bet (having the larger probability of winning) was ranked higher than the $-Bet (having the larger amount to win). Arrows pointing downward highlight pairs for which the $-Bet was ranked higher than the P-Bet.

As Fig. 2 shows, there are numerous reversals of preference across tasks. For example, people rate the gamble with a 94% chance of winning $3.00 as more attractive than the gamble with a 5% chance of winning $56.70. Rank orders are 31 versus 13. But those same people assign a *lower* selling price to the gamble with a 94% chance of winning $3.00 than to the one with a 5% chance of winning $56.70. Rank orders are 13 and 23. Furthermore, the entire pattern of preferences differs across tasks. For attractiveness ratings (A), almost all of the arrows point up; P-Bets are more attractive than $-Bets. People have risk averse preferences because, holding expected value constant, gambles with lower variance or less variability in the outcomes (P-Bets) are judged more attractive than gambles with higher variance ($-Bets). For the pricing task (C), almost all of the arrows point down; $-Bets are worth more than P-Bets. In this task, preferences are risk seeking; people assign greater worth to gambles with higher variance than to gambles with lower variance. In the strength of preference task (B), arrows point down when the probabilities of winning are small and up when the probabilities increase. Preferences are risk seeking when the probabilities of winning are small and risk averse when the probabilities increase. This pattern of preference has been well documented (Kahneman & Tversky, 1979). To summarize, preference orders are strikingly different across tasks.

Do preference orders vary when subjects are financially motivated? We ran versions of the experiments with students who were paid a flat fee to

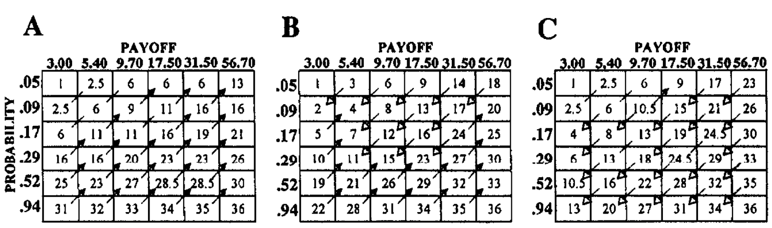

Fig. 2. Preference orders for attractiveness ratings (A), strength of preference (B), and selling price judgments (C). Larger numbers refer to more preferred gambles. Arrows show the direction of preference for pairs of gambles with equal expected values.

participate and, at the end of the experiment, were allowed to play a gamble from each task. With attractiveness ratings and selling prices, students were told that two gambles would be randomly selected from the entire set, and the gamble to which they assigned a higher rating or selling price was the gamble they would be allowed to play. With the strength of preference task, one pair was randomly selected and the gamble to which they had assigned a higher strength of preference judgment was played. Payoffs were 25% of the stated outcomes. Results were virtually identical to those shown in Fig. 2.

What happens when outcomes are losses rather than gains? We conducted another version of the experiment in which positive outcomes were converted to negative outcomes. People rated the unattractiveness of the gambles, assigned avoidance prices to the gambles (or amounts of money they would be willing to pay to avoid playing the gambles), and rated the strength of their preference for one gamble over another (i.e., the better of two evils). Preference orders in the domain of losses were the reflection of those in the domain of gains. Preferences were risk seeking with unattractiveness ratings (P-Bets were rated more unattractive than $-Bets), risk averse with avoidance prices ($-Bets had higher avoidance prices than P-Bets), and strength of preference judgments were a mixture of both risk attitudes: when the probabilities of losing were small, preferences were risk averse; but when the probabilities increased, preferences were risk seeking.

Kahneman and Tversky (1979) and others have shown that choices between risky options often reflect around the status quo. Except with small probabilities, preferences are characterized as risk averse in the gain domain and risk seeking in the loss domain. We compared preference orders for gambles based on strength of preference judgments in the gain and loss domains and found that our data also showed this pattern of reflection.

Expected values in the preceding studies ranged from $0 to $53.30 (in the gain domain) and $0 to −$53.30 (in the loss domain). These expected values are relatively small, and preference orders might differ if expected values were larger. Casey (1991) discovered that when expected values of the gambles were approximately $100, *unexpected* preference reversals occurred with choices and buying prices. That is, subjects chose the $-Bet over the P-Bet, but they assigned a higher buying price to the P-Bet than to the $-Bet.

To investigate Casey's results, we ran a version of the experiment with expected values ranging from $0.15 to $1000. Preference orders are displayed in Fig. 3; arrows connect gambles with equal expected values. Arrows point up in the attractiveness rating task (A), as found earlier (Fig. 2); P-Bets are rated more attractive than $-Bets. In the strength of preference task (B), arrows also tend to point up for all levels of probability; P-Bets

A

PROBABILITY \ PAYOFF	3.00	9.70	31.50	102	330	1070
.05	1	3	3	6	8	11.5
.09	3	5	11.5	11.5	11.5	19.5
.17	7	11.5	15.5	17	19.5	24
.29	11.5	15.5	19.5	19.5	24	24
.52	24	24	27.5	27.5	29	30.5
.94	30.5	32	33	34.5	34.5	36

B

PROBABILITY \ PAYOFF	3.00	9.70	31.50	102	330	1070
.05	1	4	7	9	15	21
.09	2	5	9	14	20	25
.17	3	8	13	19	23	28
.29	6	12	18	23	27	31
.52	9	17	25	29	32	34
.94	16	22	30	33	35	36

C

PROBABILITY \ PAYOFF	3.00	9.70	31.50	102	330	1070
.05	1	4.5	10	16	16	21
.09	2.	7	12	19	23	24
.17	3	7	14	21	26	30
.29	4.5	11	18	26	28	33
.52	7	13	21	29	31	35
.94	9	16	26	32	34	36

Fig. 3. Preference orders for attractiveness ratings (A), strength of preference (B), and buying price judgments (C) for gambles with a wider range of expected values. Arrows show the direction of preference for pairs of gambles with equal expected values.

are ranked higher than \$-Bets. Finally, in the buying price task (C), arrows tend to point down for smaller payoffs and up for larger payoffs. Thus, preference orders based on the wider range of expected values are somewhat different from those based on the smaller range, from \$0.15 to \$53.30, for strength of preference judgments and pricing judgments.

In the lower right-hand corner of each table, expected values are approximately \$100 or greater. In this region, we find no evidence of unexpected reversals. The preference order for these gambles does not seem to differ notably from those in the rest of the table for both strength of preference judgments and buying prices. If anything, there are fewer preference reversals. We find preference orders for gambles with larger expected values to be similar to those with smaller expected values, but larger stakes appear to produce more risk averse preferences in both strength of preference and pricing tasks. There is no evidence of unexpected reversals.

IV. Violations of Strong Transitivity

Preference reversals are sometimes interpreted as violations of transitivity. This interpretation, although incorrect, does not exclude the possibility that transitivity violations could occur in choices. Research on violations of strong stochastic transitivity is summarized next.

Preference reversals often involve choices, and there is an extensive literature on theories of probabilistic choice. Tests of theories are often based on stochastic transitivity (Coombs, 1983). In binary choice, a stochastic preference for gamble i over gamble j is said to occur when $P(i,j)$, the proportion of times i is chosen over j, exceeds .5. According to weak stochastic transitivity, if $P(i,j) \geq .5$ and $P(j,k) \geq .5$, then $P(i,k) \geq .5$. Moderate stochastic transitivity requires that $P(i,k) \geq \min\ [(P(i,j), P(j,k)]$, and strong stochastic transitivity states that $P(i,k) \geq \max\ [(P(i,j), P(j,k)]$.

Strength of preference judgments can also be used to test properties of transitivity. On any trial, subjects are asked to state not only the direction of their preference but also the magnitude. Assume that a judgment of 0 represents indifference on a strength of preference scale, and $S(i,j) \geq 0$ represents a judged preference for gamble i over gamble j. Suppose that $S(i,j) \geq 0$ and $S(j,k) \geq 0$. The deterministic forms of transitivity are easy to specify. Weak transitivity states that $S(i,k) \geq 0$, moderate transitivity implies that $S(i,k) \geq \min [S(i,j), S(j,k)]$, and strong transitivity states that $S(i,k) \geq \max [S(i,j), S(j,k)]$.

Empirical investigations with both animals and humans indicate that weak and moderate stochastic transitivity are often satisfied, with a few notable exceptions (Tversky, 1969). However, strong stochastic transitivity is frequently violated (Busemeyer, 1985; Rumelhart & Greeno, 1971; Tversky & Russo, 1969; Becker, DeGroot, & Marschak, 1963; Krantz, 1967; Sjöberg, 1975, 1977; Sjöberg & Capozza, 1975; Battilio, Kagel, & MacDonald, 1985). Empirical tests of these properties in strength of preference judgments also show that weak and moderate transitivities are usually satisfied, but strong transitivity often fails (Mellers & Biagini, 1994; Mellers, Chang, Birnbaum, & Ordóñez, 1992). In the experiments described previously, the median percentage of weak transitivity violations over individuals was 5%. The median percentage of moderate transitivity violations was approximately 20%. Strong transitivity violations were much more frequent, ranging from approximately 50% to 60%. This pattern occurred in choices involving both gains and losses, with large and small stakes, regardless of whether subjects were financially motivated.

Is there a pattern to the triplets that violates strong transitivity? Violations tended to occur in triplets that had at least one gamble pair with similar levels of probabilities or payoffs. When levels on one dimension were similar, differences in levels on the other dimension seemed to be enhanced, causing violations of strong transitivity. To illustrate, consider the following three gambles with equal expected values. Gamble a has a 52% chance of winning \$3.00, otherwise \$0; gamble b has a 5% chance of winning \$56.70, otherwise \$0; and gamble c has a 9% chance of winning \$17.50, otherwise \$0. With these gambles, people prefer a to b, b to c, and a to c, thus satisfying weak transitivity. The median strength of preference judgment for a over b was 10 (on a scale from 0, representing indifference, to 80, representing a very strong preference), and the median strength of preference judgment for b over c was 22. To satisfy strong transitivity, the strength of preference judgment for a over c should have exceeded 22, but it was only 12. The majority of subjects displayed this pattern in their judgments. Either the preference for gamble b over c was too large, or the preference for a over c was too small, according to strong transitivity.

In this example, gambles *b* and *c* have similar levels of probability, but different levels of payoffs. Probabilities are .05 and .09, respectively, and payoffs are $56.70 and $17.50, respectively. The similarity of probabilities may enhance differences in payoffs, making the strength of preference judgment for *b* over *c* too large, relative to the other judgments. Similar patterns are found in other triplets that violate strong transitivity.

V. Psychological Theories of Preference Reversals

Results from the preceding experiments show systematically different orders, depending on the procedure used to elicit preferences. The challenge is to discover what, if anything, remains invariant across tasks. An additional constraint on any theory of preference reversals that involves choices is that it must be able to predict the violations of strong transitivity.

We begin our discussion of invariance by representing judgment as a composition of functions, as shown in Fig. 4. In this framework, attractiveness ratings and pricing judgments can be decomposed into three processes, represented by the functions H, C, and J. Physical values of the stimuli—objective probabilities and payoffs—are represented as subjective values. Objective probabilities are transformed into subjective probabilities, and payoffs are converted to utilities. The functions that relate subjective values to physical values are H_P and H_A for probabilities and amounts, respectively.

As Weber, Goldstein, and Barlas (this volume) point out, much of psychology concerns itself with how physical values are mapped onto subjective values, and this concern dates back to research from the 1800s. As early as 1860, Fechner, a pioneer in the field of psychophysics, proposed that the physical world was related to the subjective world by means of a logarithmic relationship. Furthermore, in his classic book, Elements of Psychophysics, Fechner (1860/1966) argued that the logarithmic function had been proposed as the link between physical and subjective values 100 years earlier by Daniel Bernoulli, who used it to describe the relationship between *fortune morale* (moral wealth) and *fortune physique* (physical wealth). Although there is still controversy about the shape of the H function in a variety of different tasks, there is no disagreement about the need to postulate subjective values.

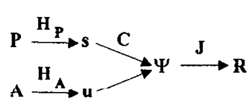

Fig. 4. Framework for analysis of judgment.

Subjective values of probabilities and payoffs, denoted s and u, are combined by means of a function, C, to form a subjective impression of the gamble, represented as Ψ. When the dependent variable is an attractiveness rating, Ψ represents the composite attractiveness of the gamble. When the dependent variable is a pricing judgment, Ψ represents the composite worth of the gamble. These subjective impressions are then transformed to numerical responses, R, such as a number on a category rating scale or a dollar amount, by means of a judgment function, J. This function represents the response transformation and is typically assumed to be monotonic.

First, we consider three psychological theories of preference reversals: contingent-weighting theory (Tversky et al., 1988), reference-level theory (Luce, Mellers, & Chang, 1992), and expression theory (Goldstein & Einhorn, 1987). Then, we discuss a fourth possibility, the change-of-process theory, in more detail.

A. Contingent-Weighting Theory

Tversky et al. (1988) proposed a hierarchy of contingent trade-off models to account for discrepancies between judgment and choice. A special case is the contingent-weighting model, which Tversky et al. proposed to account for preference reversals between buying price judgments and attractiveness ratings of two-outcome gambles with some probability p of winning amount A, otherwise winning nothing. According to this model, the weights associated with the attributes depend on their compatibility with the response scale. Tversky et al. assumed that the value of a gamble is the product of probability and payoff, with each attribute weighted by a power function that depends on its relationship to the response scale. Using additive, linear multiple regression on the logarithms of the responses, they argued that the relative weight for payoffs is greater in the pricing task than in the rating task because, in a pricing task, the independent variable (payoffs in terms of dollars) is more compatible with the response scale (dollars).

According to this account, preference reversals are caused by changes in the weights of the attributes. Differential weighting might be attributed to changes in the C function; however, the aggregation rule does not vary. Furthermore, Tversky et al. did not have independent estimates of weights and scales. For these reasons, we prefer to treat contingent weighting as a theory that attributes preference reversals to changes in the H function. Tversky et al. (1988) presented a more general contingent trade-off model in which both scales and processes can vary with the task, but this more general model makes no predictions about when scales or processes should vary.

B. Reference-Level Theory

Luce et al., (1992) proposed a theory of preference reversals for certainty equivalents (CEs) and choices. Certainty equivalents (in this case, selling prices) are assumed to be the basic primitive, and choices are derived from comparisons of CEs. When confronted with a pair of gambles and asked to make a choice, people assess the monetary worth, or CE, of each gamble relative to a reference level that is unique to each pair of gambles. If either of the two CEs is positive, the reference level is the smaller of the two CEs. If both of the CEs are losses, the reference level is the smallest CE. Before making a choice, subjects are assumed to recode the outcomes of the gambles relative to the reference level. This recoding reflects the fact that even gains can feel like losses when expectations are higher than positive outcomes. Finally, gambles are evaluated by means of their rank- and sign-dependent utility (Luce, 1991; Luce & Fishburn, 1991). In this account, preference reversals also occur as the result of changes in the H function that vary systematically for each pair of gambles.

C. Expression Theory

Goldstein and Einhorn (1987) proposed another account, called *expression theory,* in which preference reversals occur in the J function. Although this function is typically assumed to be monotonic, Goldstein and Einhorn postulated a nonmonotonic function that differs systematically for each gamble. This transformation presents gambles in terms of their "proportional worth." The proportional worth of the gamble on a pricing scale is defined with respect to the gamble's best and the worst outcomes. Similarly, the proportional worth of the gamble on a rating scale is defined with respect to the highest and lowest permissible ratings. Subjects convert gambles to their proportional worths, and those proportional worths should be monotonically related across tasks.

Mellers, Ordóñez, and Birnbaum (1992) present evidence inconsistent with contingent-weighting theory and expression theory, and Luce et al. (1992) discuss the pros and cons of reference-level theory. Additional discussions of these theories can be found elsewhere. Now, a fourth theory, referred to as *change-of-process* theory will be presented in more detail.

D. Change-of-Process Theory

Mellers, Chang, Birnbaum, and Ordóñez (1992) and Mellers, Ordóñez, and Birnbaum (1992) have suggested using a change-of-process theory to describe preference reversals. In this account, decision strategies by which subjects combine information vary across tasks and subjective values of the payoffs remain constant. That is, the combination rule (C in Fig. 4) depends on the response mode, but the psychophysical function (H) remains constant. A number of researchers (Payne, 1982; Payne, Bettman, & Johnson, 1992, 1993; Smith, Mitchell, & Beach, 1982) discuss the possibility that decision makers use different decision strategies depending on the effort required, the accuracy needed, and the cost of decision errors. Some have theorized that preference reversals should be attributed to different decision strategies across tasks (Lichtenstein & Slovic, 1971; Schkade & Johnson, 1989; Johnson, Payne, & Bettman, 1988). Change-of-process theory extends these ideas by postulating specific models for ratings, prices, and strength of preference judgments. In the strength of preference judgments, weights depend not on compatibility but rather on the similarity of the attribute levels under consideration. Furthermore, the change-of-process theory has the additional assumption of utility invariance; that is, the utilities associated with monetary outcomes are assumed to be invariant across response modes (Birnbaum, 1974; Birnbaum & Veit, 1974).

The importance of this assumption cannot be overstated. Utility invariance has theoretical consequences because it implies that preference measurements have meaning over and beyond the task in which they are derived. This constraint increases the psychological validity and usefulness of the measurements and helps create a more solid theoretical framework in which both models and measurements contribute to our understanding of preference reversals.

Consider gamble *a*, with some probability *p* of winning an amount *x* or otherwise winning nothing. The change-of-process theory asserts that under some conditions, the attractiveness rating of gamble *a* can be described by an additive combination of utility and subjective probability, and scales are equated by means of a constant, k, as follows:

$$A(a) = J_A[k\ s_a + u_a], \tag{1}$$

where A is the attractiveness rating of gamble *a*. J_A is a monotonic response function, k is the scaling constant, s_a is the subjective probability associated with probability p, and u_a is the utility associated with amount *x*. The selling price for gamble *a* is described by a multiplicative model:

$$P(a) = J_P[s_a\ u_a], \tag{2}$$

where P is the selling price of gamble *a*, J_P is the response function, and s_a and u_a are the subjective probability and utility, respectively. Finally, strength of preference judgments are described by both operations. That is, the strength of preference for one gamble over another is a monotonic function of the difference between the two subjective products, with one additional assumption: contrast weighting. This assumption, which is necessary to account for violations of strong transitivity, is that people focus attention primarily on attributes that differ. Utilities and subjective probabilities are weighted according to the contrast between levels along a dimension. For example, if gambles have similar probabilities of winning, the weight given to the probability dimension is smaller than if probabilities differ. Likewise, if gambles have similar payoffs, the weight given to the payoff dimension is smaller than if payoffs differ. Contrast weights may represent an attentional focus (Nosofsky, 1986). The discounting of dimensions that are not terribly diagnostic may be one way to simplify the task and minimize effort (Payne, 1976; Payne et al., 1992).

The contrast-weighting model is expressed as follows:

$$S(a,b) = J_S[u_a^{\alpha}\, s_a^{\beta} - u_b^{\alpha}\, s_b^{\beta}], \tag{3}$$

where $S(a,b)$ is the strength of preference for gamble a over gamble b, J_S is the judgment function, α is the contrast weight for utilities, β is the contrast weight for probabilities, and the other symbols are as defined earlier. To fit this model to any set of data, it is necessary to define the terms *similar* and *dissimilar.* Mellers, Chang, Birnbaum, and Ordóñez (1992) used a simple proxy for similarity; adjacent levels of probability and payoffs in the experimental design (e.g., \$3.00 and \$5.40, \$5.40, and \$9.70 in Fig. 1) were treated as similar and received one contrast weight, and nonadjacent levels were treated as dissimilar and received a different contrast weight. Thus, two contrast weights were allowed for each dimension, one when contrasts were small and the other when contrasts were large.

This proxy for similarity is undoubtedly too simple; Goldstone, Medin, and Gentner (1991) show that similarity is often relational. That is, the degree to which a shared feature affects similarity depends on the other features shared by those objects. Nonetheless, the proxy used by Mellers, Chang, Birnbaum, and Ordóñez (1992) worked well at describing the patterns of strong transitivity violations described earlier. That is, the similarity of levels on one attribute intensifies differences on other attributes.

Fits of the change-of-process theory to the experiments described earlier have been quite promising. The theory can describe the changing preference orders with little residual variance. Estimated contrast weights for similar levels along a dimension are smaller than weights for dissimilar levels. Furthermore, the theory accurately predicts the violations of strong transitivity in strength of preference judgments. Finally, the assumption of stable utilities is satisfied. That is, utilities are invariant despite labile preferences.

In sum, change-of-process theory gives an excellent description of judgments and choices in the domain of risky decision making.

VI. Deciding How to Decide

If subjects actually use different strategies to evaluate gambles, what factors determine strategy selection for a particular task? One factor that appears to influence the choice of a decision strategy is the stimulus context. We have found that when subjects rate the attractiveness of gambles, attractiveness ratings appear parallel, consistent with the interpretation that subjects combine probabilities and payoffs additively. However, the additive model makes some implausible predictions. According to this model, payoffs and probabilities should make independent contributions to the overall attractiveness of the gamble. If subjects are shown gambles with varying probabilities of winning $0, the attractiveness of the gamble should *increase* as the probability of winning *increases*. Similarly, if subjects are shown a set of gambles with a 0% chance of winning various amounts, the attractiveness of the gambles should increase as the amount to win increases.

We examined whether people show this peculiar behavior by presenting them with a set of gambles, some of which had zero probabilities of winning various amounts or varying probabilities of winning $0. For the majority of subjects, attractiveness ratings were no longer parallel. Instead, ratings showed a bilinear interaction consistent with a multiplicative combination rule of probability and payoff. It is likely that the inclusion of these unusual gambles highlighted the problems of an additive strategy, and therefore subjects switched to a multiplicative strategy. The process by which subjects combine information may depend not only on the response mode, but also on the surrounding stimulus context.

Payne et al. (1992, 1993) argued that people have a repertoire of decision strategies from which they select a particular strategy, depending on the task, the context, the cost of decision errors, the accuracy required, and the effort required. This framework implies that preference reversals will diminish when the cost of a decision error increases or effort decreases. Most tests of this hypothesis have focused on the cost of decisions. Studies that build in financial incentives increase the costs of decision errors, yet preference reversal rates are often unaffected by financial incentives, a finding that runs counter to the effort–accuracy framework for selecting strategies.

Johnson et al. (1988) examined the level of effort required of the decision maker by manipulating the display of probabilities. They reasoned that simple probability displays should produce fewer preference reversals than more complex displays; and, in fact, they found a somewhat lower rate of preference reversals when probabilities were presented in a simple form (e.g., .88 or 7/8) compared with a more complex form (399/456). They argued that people expend less effort and can more easily perform expected-value computations with simple displays than with more complex displays. Johnson et al. reasoned that if subjects use expected-value computations, their preferences would be more consistent across tasks.

However, in another study, González-Vallejo and Wallsten (1992) varied the display of probabilities by using numerical and verbal forms and found that preference reversals were significantly reduced with verbal probabilities. It seems unlikely that verbal probabilities encouraged expected-value computations. Thus, the evidence supporting the effort–accuracy framework for preference reversals appears mixed. As Payne, Bettman, and Johnson point out, many factors presumably influence the decision about how to decide, and the effects of these factors do not appear to be simple.

VII. Which Preference Order Represents "True" Preferences?

The change-of-process theory asserts that people use different decision strategies for combining information, and those decision strategies produce different preference orders. Does one of those preference orders better reflect people's "true" preferences? This question presupposes a set of "true" preferences, a somewhat controversial assumption. Perhaps an easier question to answer is whether any of the preference orders derived from pricing judgments, attractiveness ratings, or strength of preference judgments is more stable and robust than the others. Ordóñez, Mellers, Chang, and Roberts (1994) pointed out that a common thread runs through almost all of the past research on preference reversals: subjects perform the tasks *sequentially.* That is, they evaluate a set of gambles with one response mode and later reevaluate the same set with a second response mode. By using this procedure, subjects are never directly confronted with their own inconsistencies.

We attempted to make subjects aware of their preference reversals by asking them to perform two tasks *simultaneously.* In this case, they evaluated pairs of gambles with both response modes before continuing to the next pair of gambles. Pairs of response modes were selected from three possible tasks (attractiveness ratings, selling prices, and strength of preference judgments). We thought that this procedure might motivate subjects to produce more consistent responses. We provided additional motivation for consistency by telling subjects that at the end of the experiment we would randomly select a trial that contained a pair of gambles. If their responses in the two tasks were consistent (e.g., they priced gamble *a* higher than gamble *b* and they also rated gamble *a* as more attractive than gamble *b*), they would be allowed to play the preferred gamble. If their responses were inconsistent, the experimenter would select the gamble to be played, and that gamble might not be the one they preferred.

If this procedure yields a single preference order across tasks, there are at least three different ways in which preferences could become more consistent. First, preference orders for gamble pairs could be determined on the basis of the first task performed; the preference order for the second task would always be identical to the preference order for the first task. Second, the preference order for one of the two tasks could dominate the preference order for the other task. For example, if selling prices dominated the other tasks, the preference orders for the other tasks would resemble that of selling prices, regardless of task order. Third, preference orders from the two tasks could merge into a new preference order. This order presumably would be some composite of the original two preference orders.

When financially motivated subjects made responses to the two tasks simultaneously, preference reversals were not only reduced but they were actually eliminated for two of the three task pairs. The derived preference order for each task appeared to merge into a new, compromise order that differed depending on the pair of tasks. For attractiveness ratings and selling prices, the merging was incomplete (i.e., the two tasks still had somewhat different preference orders) and reversals continued to occur.

Does one of these preference orders represent subjects' true preferences? No single preference order appeared across the relevant task combinations. It seems unlikely that a true preference order exists, as subjects were presumably just responding to the procedure and incentives in the experiment. They appeared to construct their preferences as some compromise of the preference orders from each separate task. Another procedure and incentive structure might have produced a different set of consistent preference orders. Results from this study suggest that subjects can give consistent responses when motivated to do so, but these preferences do not necessarily reflect true preferences, and the concept of "true preference" is, itself, somewhat dubious.

VIII. Preference Reversals in Riskless Domains

Goldstein and Weber (this volume) argue that decision researchers have been excessively drawn to the gambling metaphor. It is easy to see why this seductive tool has been so widely used. It contains the main elements of a risky decision, namely, alternative actions controlled by the decision maker, states of nature that constitute the environment, and outcomes that are the results of actions and states of nature. Furthermore, with gambles, these elements are easy to manipulate. Goldstein and Weber argue that because of this simplicity, the central assumption of the gambling metaphor, that decisions depend on some combination of degree of belief and degree of desirability, may not generalize to other types of decisions. If so, our theories of decisions based on gambles may not generalize to nongambling contexts. Therefore, it seems reasonable to ask whether the change-of-process theory can also describe preference reversals outside the domain of gambles. Do people use different decision strategies in riskless domains?

To answer this question, Mellers and Cooke (1994) examined preference reversals using apartments described by monthly rent and distance to campus. One group of students rated the attractiveness of apartments, and another group made choices between pairs of apartments. Figure 5 shows the preference orders for the two tasks. Arrows connect pairs of nondominated apartments that differ in rent and distance; one apartment is closer to campus, and the other apartment is cheaper. Downward arrows represent preferences for closer apartments over cheaper apartments. Upward arrows show preferences for cheaper apartments over closer apartments. In the rating task (A), most of the arrows point down: closer apartments are more attractive than cheaper apartments. In the choice task (B), all of the arrows point up: cheaper apartments are chosen over closer apartments.

Why might these reversals occur? It seems quite reasonable that students would rate closer apartments as more attractive. Closer apartments are usually more convenient. However, if closer apartments are more attractive, why are they not chosen? Perhaps students want to save their money; they prefer to spend less. But if they want to save money, then why do they not rate cheaper apartments as more attractive? These two tasks may highlight different features of the apartments: one task may represent preferences under no constraints and the other may represent preferences with realistic

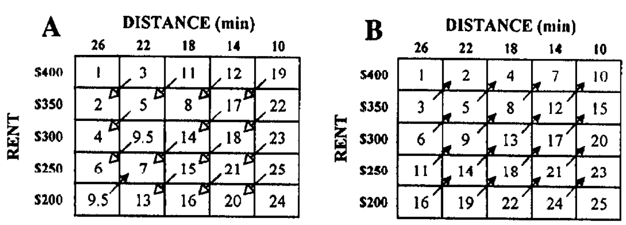

Fig. 5. Preference orders for attractiveness ratings (A) and choices (B) for apartments described by monthly rent and distance to campus. Arrows show the direction of preference for pairs of apartments differing in rent and distance.

aspirations. However, nothing in the task instructions guided subjects to take different points of view.

Mellers and Cooke (1994) did *not* find evidence to support the change-of-process theory. Both attractiveness ratings and choices were consistent with a single decision strategy. If changes in decision strategies cannot account for preference reversals, then what can?

Tversky et al. (1988) proposed two hypotheses to describe why subjects weight attributes differently across tasks. *Compatibility,* described earlier, focuses on the similarity of the attribute and response scales. The second hypothesis, *prominence,* states that the more prominent attribute weighs more heavily in choices than in matching tasks. People tend to make choices according to the more important dimension(s), but they match options by comparing trade-offs along two or more dimensions. If rent is more prominent than distance in evaluations of apartments, rent will receive relatively more weight in the choice task than in the rating task.

Although Tversky et al. (1988) proposed the prominence hypothesis to account for discrepancies between choice and matching tasks, this hypothesis can also be tested for choice and rating tasks. If people assign greater weight to the more important attribute in choice tasks than in rating tasks, one would expect preference reversals to be different for people who differ in their opinion about the more important attribute. In particular, people who say that distance is more important might choose closer apartments over cheaper apartments, but rate cheaper apartments as more attractive. Those who say that rent is more important might choose cheaper apartments over closer apartments, but rate closer apartments as more attractive.

To test this hypothesis, Mellers and Cooke (1994) investigated preferences for apartments using a within-subject design in which students assigned ratings to apartments and chose between pairs of apartments. After completing each task, they indicated which attribute, rent or distance, was more important by allocating 100 points between the two attributes to reflect their relative importance. A few individuals reversed their relative importance judgments for the two tasks. A few others said the attributes were equally important in one or both tasks. The remaining students fell into two groups, those who said that rent was more important than distance and those who said that distance was more important than rent.

Figure 6 shows the patterns of preference for the two groups. Comparisons were made over both subjects and apartment pairs. Entries are percentages from 4900 comparisons in panel A and 6000 comparisons in panel B. Panel A shows results for people who said that rent was more important than distance. These people chose cheaper apartments more frequently than closer apartments in 66% of the pairs (18% + 48%). They also rated cheaper apartments as more attractive than closer apartments in 56% of

A

RATINGS	CHOICES: Closer	CHOICES: Cheaper
Closer	27	18
Cheaper	8	48

B

RATINGS	CHOICES: Closer	CHOICES: Cheaper
Closer	66	9
Cheaper	10	15

Fig. 6. Patterns of preference for subjects who assigned ratings and made choices between pairs of apartments. Preferences are shown for those who say that rent is more important than distance (A) and for those who say that distance is more important than rent (B).

the pairs. Panel B shows data for people who said that distance was more important than rent. In 76% of the pairs, these subjects chose closer apartments more often than cheaper apartments, and in 75% of the pairs, they rated closer apartments as more attractive. Thus, preferences in the two groups differed in the expected directions, consistent with their judged importance weights.

What about preference reversals? Students who thought rent was more important picked cheaper apartments over closer apartments and rated closer apartments as more attractive significantly more often than the opposite preference reversal (18% vs. 8%). On the other hand, students who thought distance was more important chose the closer apartment in the choice task, and rated the cheaper apartment as more attractive slightly more often than the opposite reversal (10% vs. 9%). Although these latter two preference reversals are very similar, they are in the direction predicted by the prominence hypothesis, suggesting that preference reversals depend on which attribute is judged to be more important.

If preference reversals interact with judged importance weights, students in both groups who have extreme judged weights should show even stronger preference reversals. Results were consistent with this prediction; subjects with more extreme judged weights demonstrated more pronounced reversals across choices and ratings. For example, those who assigned 75 points or more to rent chose the cheaper apartment, but rated the closer apartment as more attractive 22% of the time. The opposite reversal occurred only 4% of the time. Subjects who assigned 75 points or more to distance chose the closer apartment, but rated the cheaper apartment as more attractive 11% of the time, and the opposite reversal occurred in only 4% of the cases. These results suggest that, at least in some cases, the prominence hypothesis can account for differing orders in choice and rating tasks, and the attribute judged more important is weighted more heavily in choices.

To investigate other similarities between risky and riskless domains, we constructed 100 triplets of the form $p(a,b)$, $p(b,c)$, and $p(a,c)$ from apartment choice pairs in the within-subject experiment. For both groups, weak stochastic transitivity was always satisfied, moderate stochastic transitivity was almost always satisfied, and violations occurred in only 8% of the triplets for both groups. However, strong stochastic transitivity failed in more than one-third of the triplets; violations occurred in 36% and 37% of the triplets for the groups who said that distance was more important and rent was more important, respectively.

To account for preference reversals and violations of strong stochastic transitivity, we proposed that (1) subjects average information about rent and distance in both tasks, (2) attribute weights can vary with tasks and subjects assign greater weight to the more important dimension in choices than in ratings, (3) weights in the choice task also vary depending on the similarity of levels along an attribute, and (4) utilities remain invariant across tasks. We will now show how each of these assumptions is incorporated into a change-of-weight theory that formalizes the prominence hypothesis for a riskless domain, applies it to choices and ratings, and adds two important new assumptions—contrast weighting and utility invariance.

Averaging. In the rating task, subjects are assumed to average attribute information about monthly rent and distance as follows:

$$\mathrm{R}(a) = \mathrm{J_R}[\mathrm{w}\ \mathrm{r}_a + (1 - \mathrm{w})\ \mathrm{d}_a], \tag{4}$$

where $\mathrm{R}(a)$ is the attractiveness rating of apartment a, $\mathrm{J_R}$ is a linear function, w is the weight of rent, r_a is the utility of rent, and d_a is the utility of distance. In the choice task, attributes are also averaged and then compared by means of a subtractive operation:

$$\mathrm{P}(a,b) = \mathrm{J_P}[(\mathrm{w}^*\ \mathrm{r}_a + (1 - \mathrm{w}^*)\ \mathrm{d}_a) - (\mathrm{w}^*\ \mathrm{r}_b + (1 - \mathrm{w}^*)\mathrm{d}_b)], \tag{5}$$

where $\mathrm{P}(a,b)$ is the proportion of subjects who choose apartment a over apartment b, $\mathrm{J_P}$ is a logistic function with one slope parameter, w* is the weight of rent in the choice task that varies depending on the similarity of rents for apartments a and b, r_a and r_b are the rent utilities for apartments a and b, and d_a and d_b are the distance utilities for apartments a and b.

Change-of-weight. Weights in Eq. 4 and Eq. 5 are represented as

$$\mathrm{w} = \mathrm{w_r} / (\mathrm{w_r} + \mathrm{w_d}) \tag{6}$$

and

$$w^* = w_r^* / (w_r^* + w_d^*) \quad (7)$$

in the rating and choice tasks, respectively. These weights are allowed to differ across tasks. Subjects assign greater weight to the more important dimension in choices than in ratings. If rent is judged to be the more important dimension, then w_r^* is greater than w_r, and if distance is judged to be more important, w_d^* is greater than w_d.

Contrast weighting. Attribute weights in choice are allowed to vary with the similarity of attribute levels. Thus, in the choice task, there are two sets of weights for each attribute (or two values for each w_r^* and w_d^*), one for when attribute levels are similar (levels that are adjacent in the experimental design), and one for when attribute levels are dissimilar.

Utility invariance. The utilities of rent and distance, r and d, are assumed to be identical in the two tasks (Eqs. 4 and 5).

We fit this change-of-weight theory to mean attractiveness ratings and choice proportions separately for each group of subjects. Predicted preference reversals closely resembled the observed orderings. In addition, estimated weights of rent and distance were in the predicted direction. The estimated weight of rent for subjects who said that rent was more important than distance was greater in the choice task than in the rating task. The estimated weight of distance for those who said that distance was more important than rent was also greater in the choice task than in the rating task. Furthermore, contrast weights were in the predicted direction; similar levels along an attribute received less weight than dissimilar levels. Finally, the theory described the observed violations of transitivity in choice proportions.

In summary, the change-of-weight theory provides a good account of preference reversals in a riskless domain. The direction of preference reversals systematically differs depending on the attribute judged more important. Rent has a greater effect on choices than on ratings when rent is judged more important, and distance has a greater effect on choices than ratings when distance is judged more important. Furthermore, violations of strong stochastic transitivity can be described by contrast weighting. When two apartments have similar levels along an attribute, that attribute receives less weight than when levels differ. These weight changes allow the model to capture both preference reversals and violations of strong stochastic transitivity. Last, but not least, all of this is accomplished while maintaining the assumption of utility invariance. The utilities of rent and distance are invariant across tasks. Thus, even though preferences are labile, measurements of utilities need not be. With the proper model of preference reversals, utilities remain stable and have meaning beyond the task from which they were derived.

IX. A Two-Pronged Explanation of Preference Reversals

Preference reversals do not *always* occur. There are many instances in which we are perfectly clear about what we want, regardless of how the question is asked. These situations are ones in which one option dominates another, one option has a decisive advantage, or when conflicts between attribute trade-offs have been resolved. However, preference measurement is much more complicated when trade-offs are difficult, when uncertainties are hard to specify, and when our preferences themselves are unstable (March, 1978). In these cases, preferences can be influenced by a variety of seemingly irrelevant factors (Payne et al., 1992).

Why do preferences reverse across response modes? Results from the preceding studies in risky and riskless domains suggest that they reverse for at least two reasons: (1) subjects appear to change either strategies or weights across tasks, and (2) they appear to change decision strategies in risky domains and attribute weights in riskless domains. Why might the introduction of risk change the decision strategies used in ratings, prices, and choices? In risky alternatives, probability is an attribute, and probabilities may be conducive to different interpretations. In attractiveness ratings of monetary lotteries, probability is often treated as an independent variable that contributes to the overall attractiveness of the lottery in an additive fashion. For pricing judgments, however, probability is treated as a discount factor, moderating outcome information in a multiplicative fashion. Other attributes (e.g., price or rent) do not have similar interpretative ambiguity. However, even though different mechanisms appear to be responsible for preference reversals, stable measures of utility can be found across a variety of different tasks.

Several studies have shown that it *is* feasible to maintain utility invariance across tasks. Birnbaum, Coffey, Mellers, and Weiss (1992) accounted for differences between buying and selling prices of risky options by assuming that buyers and sellers have identical utilities for monetary payoffs but differ in the decision weights associated with outcomes. Buyers tended to be more pessimistic and weighted lower-valued outcomes more heavily, whereas sellers tended to be more optimistic and weighted higher-valued outcomes more heavily. The question of whether utilities vary in risky versus riskless decisions has been a longstanding controversy (see Bell & Raiffa, 1988; Dyer & Sarin, 1982; von Winterfeldt & Edwards, 1986). Birnbaum and Sutton (1992) managed to preserve utility invariance for monetary payoffs across risky and riskless domains by assuming that differences between risky and riskless choices were the result of different decision weights that depended on the rank order of the outcome, with lower-valued outcomes being weighted more heavily than higher-valued outcomes. Finally, Weber, Anderson, and Birnbaum (1992) described attractiveness ratings and risk ratings for the same gambles by assuming that the utilities of the payoffs were identical across the two tasks. Differences between tasks, and even differences among individuals, were explained solely by changes in decision weights.

Although there are still many unanswered questions, the results from these studies suggest that preference reversals can occur from either changes in weight or changes in process. Understanding the mechanisms that produce both types of changes and investigating how they are affected by memory, attention, and emotional states are fruitful areas for future research. In the meantime, it is encouraging and of considerable practical usefulness that we can obtain stable measures of utility despite labile preferences. With appropriate models of judgment and choice, we *can* estimate utilities that do not depend on the method used to infer preference.

Acknowledgments

This research was supported by National Science Foundation Grant SES-9023160 to B. A. Mellers, National Science Foundation Grant SES-9022192 to E. U. Weber with financial support from the Graduate School of Business, University of Chicago, and a Ford Foundation Fellowship to L. D. Ordóñez. Preparation of the manuscript occurred while the E. U. Weber was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA. The authors thank Jerry Busemeyer and Doug Medin for helpful comments.

References

Battilio, R. C., Kagel, J. H., MacDonald, D. N. (1985). Animals' choices over uncertain outcomes: Some initial experimental results. *American Economic Review, 75,* 597–613.

Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. Unpublished manuscript. Northwestern University.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science, 8,* 41–55.

Bell, D. E., & Raiffa, H. (1988). Marginal value and intrinsic risk aversion. In D. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions.* New York: Wiley.

Berg, J. E., Dickhaut, J. W., & O'Brien, J. R. (1985). Preference reversal and arbitrage. In V. Smith (Ed.), *Research in Experimental Economics* (Vol. 3, pp. 31–32). Greenwich, CT: JAI Press.

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

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., & 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,* 331–346.

Birnbaum, M. H., & Veit, C. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception and Psychophysics, 15,* 7–15.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sample models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 538–564.

Casey, J. (1991). Reversal of the preference reversal phenomenon. *Organizational Behavior and Human Decision Processes, 48,* 224–251.

Chapman, G. B., & Johnson, E. J. (1993). Preference reversals in monetary and life expectancy evaluations. Unpublished manuscript. University of Illinois.

Coombs, C. (1983). *Psychology and mathematics.* Ann Arbor: University of Michigan Press.

Cox, J., & Grether, D. (1992). *The preference reversal phenomenon: Response mode, markets, and incentives.* Social Science Working Paper 810. California Institute of Technology.

Chu, Y., & Chu, R. (1990). The subsidence of preference reversals in simplified and marketlike experimental settings. A note. *The American Economic Review, 80,* 902–911.

Dyer, J. S., & Sarin, R. K. (1982). Relative risk aversion. *Management Science, 28,* 875–886.

Fechner, G. (1966). *Elements of psychophysics.* D. H. Howes & E. G. Boring (Eds.). H. E. Alder (Trans.). New York: Holt, Rinehart, and Wilson. (Original work published 1860)

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

Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. R. Busemeyer, R. Hastie, D. L. Medin (Eds.), *The Psychology of Learning and Motivation,* this volume.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23,* 222–262.

González-Vallejo, C., & Wallsten, T. (1992). Effects of probability mode on preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 855–864.

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

Irwin, J. R., Slovic, P., Lichtenstein, S., & McClelland, G. H. (1993). Preference reversals and the measurement of environmental values. *Journal of Risk and Uncertainty, 6,* 5–18.

Johnson, E. J., Payne, J. W., & Bettman, J. R. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes, 42,* 1–21.

Krantz, D. (1967). Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology, 4,* 226–245.

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

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

Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology, 101,* 16–20.

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

Luce, R. D. (1991). Rank- and sign-dependent linear utility models for binary gambles. *Journal of Economic Theory, 53,* 75–100.

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., Mellers, B. A., & Chang, S. (1992). 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.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics, 9,* 587–608.

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

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

Mellers, B. A., & Cooke, A. D. J. (in press). The role and task and context in preference measurement. Psychological Science.

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

Mowen, J. C., & Gentry, J. W. (1980). Investigation of the preference reversal phenomenon in a new product introduction task. *Journal of Applied Psychology, 65,* 715–722.

Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Ordóñez, L., Mellers, B., Chang, S., & Roberts, J. (in press). Are preference reversals reduced when made explicit? Organizational Behavior and Human Decision Processes.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92,* 382–402.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology, 43,* 87–131.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Pommerehne, W. W., Schneider, F., & Zweifel, P. (1982). Economic theory of choice and the preference reversal phenomenon: A reexamination. *American Economic Review, 72,* 569–574.

Reilly, R. J. (1982). Preference reversal: Further evidence and some suggested modifications in experimental design. *American Economic Review, 72,* 576–584.

Rumelhart, D., & Greeno, J. (1971). Similarity between stimuli: An experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology, 8,* 370–381.

Schkade, D. A., & Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes, 44,* 203–231.

Sjöberg, L. (1975). Uncertainty of comparative judgments and multidimensional structure. *Multivariate Behavioral Research, 10,* 207–218.

Sjöberg, L. (1977). Choice frequency and similarity. *Scandinavian Journal of Psychology, 18,* 103–115.

Sjöberg, L., & Capozza, D. (1975). Preference and cognitive structure of Italian political parties. *Italian Journal of Psychology, 2,* 391–402.

Smith, J. F., Mitchell, T. R., & Beach, L. R. (1982). A cost–benefit mechanism for selecting problem-solving strategies: Some extensions and empirical tests. *Organizational Behavior and Human Performance, 29,* 370–396.

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

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

Tversky, A., & Russo, J. (1969). Substitutability and similarity in binary choice. *Journal of Mathematical Psychology, 6,* 1–12.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reveral. *American Economic Review, 80,* 204–217.

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

Weber, E. U. (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.

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.

Wedell, D. H., & Böckenholt, U. (1990). Moderation of preference reversals in the long run. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 429–438.