ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
Vol. 62, No. 3, June, pp. 300–317, 1995

Preference Reversals in Monetary and Life Expectancy Evaluations

GRETCHEN B. CHAPMAN

University of Illinois at Chicago

AND

ERIC J. JOHNSON

The Wharton School, University of Pennsylvania

Two experiments demonstrate a new type of preference reversal. In life expectancy evaluations, health items (e.g., a new treatment that would give you perfect 20/20 vision) were preferred to commodities (e.g., 1 day of vacation in Bermuda), but in monetary evaluations, commodities were preferred to health items. These reversals result from the pattern of similarity of commodities and health items to money and life expectancy and are therefore an example of Tversky, Sattath, and Slovic's (1988) semantic compatibility principle.

Increasing evidence indicates that people often have preferences that are not well defined or stable across different modes of measurement. Because many preference judgments appear to be constructed on-line in the context of a specific preference elicitation procedure (e.g., Payne, Bettman, & Johnson, 1992), the revealed preferences vary both across different response elicitation tasks (for example, judgment versus choice) or even within response tasks across different response scales (for example, pricing versus rating). The lability of preference judgments can result in reversals of preference (e.g., Goldstein & Einhorn, 1987; Lichtenstein & Slovic, 1971; Lindman, 1971; Tversky, Sattath, & Slovic, 1988).

In a typical preference reversal experiment (e.g., Lichtenstein & Slovic, 1971) subjects evaluate a lottery with a small chance of winning a large monetary amount (a $-bet) and a lottery with a large probability of winning a modest cash amount (a P-bet). In choosing between the two lotteries, subjects generally prefer the P-bet; often, however, subjects also state a higher cash equivalent for the $-bet. Thus, the preference for the two lotteries varies with the response elicitation procedure, a violation of normative theory.

Tversky *et al.* (1988) offer an explanation for such preference reversals. According to their compatibility principle, the weight of an input component is enhanced by its compatibility with the output. For example, since monetary payoffs are compatible with a monetary response scale, cash evaluations of lotteries will be more influenced by the payoffs (relative to the probabilities) than will choices between lotteries. This mechanism is termed contingent weighting because the weight of each input attribute is contingent on the compatibility between the attribute and the response scale.

This explanation also applies to reversals that occur when only one response task, such as judgment, is used with different response scales. For example, Slovic, Griffin, and Tversky (1990) asked subjects to predict the performance of 10 students in one course based on their grades in a second course and class rank in a third course. The class grade inputs were given more weight when performance predictions were in terms of grades. Conversely, class rank inputs were given more weight when performance predictions were made in term of class rank. As a result of these differing attribute weights, predictions from the class rank and from the class grade response modes did not agree.

In another example of a compatibility effect, Schkade and Johnson (1989, Experiment 2), replicating a result first demonstrated by Goldstein and Einhorn (1987), asked subjects to evaluate P-bets and $-bets by rating

Support for this research was provided by NSF SES90-10535 to the first author and NSF SES88-09299 to the second author. This research was presented at the November, 1993 meeting of the Society for Judgment and Decision Making. We thank Christine Emler, Macy Ng, and Wendy Park for running subjects and two anonymous reviewers for comments on a previous version of this paper. Correspondence and reprint requests should be addressed to Gretchen Chapman, University of Illinois at Chicago, Department of Medical Education (M/C 591), 808 South Wood Street, #986, Chicago, IL 60612-7309.

0749-5978/95 $6.00

each on a 100-point scale and by indicating their minimum selling price for each. Because the 100-point rating scale was compatible with probability, probabilities were most influential for that response mode; thus P-bets tended to be rated higher than $ bets. Likewise, because the monetary pricing scale was compatible with monetary payoffs, payoffs were most influential for that response mode; thus $-bets tended to be priced higher than P-bets. The differences between the two response modes resulted in preference reversals. For 70% of the lottery pairs, the P-bet was preferred in rating, and the $-bet was preferred in pricing.

While the compatibility hypothesis describes the increased weight given to compatible attributes, several specific mechanisms may produce these effects. To explain discrepancies between choice and judgment, Tversky *et al.* (1988) note that these two tasks invoke two different kinds of reasoning. Judgment requires quantitative reasoning, while choice can be accomplished by qualitative comparison. Fischer and Hawkins (1993; Hawkins, 1994) note that qualitative preference tasks are more likely than quantitative tasks to evoke a preference for the alternative that is superior with respect to the most important attribute. This principle is termed strategy compatibility. A second mechanism that can produce compatibility effects has been termed scale compatibility (Tversky *et al.*, 1988): different response scales are more or less compatible with the attributes of the item to be judged. For example Slovic *et al.* (1990) argue that a grading scale is more similar to past grades than to past class rank, causing that element of the student profile to be overweighted. Schkade and Johnson (1989) similarly suggest that pricing judgments emphasize the amount to win, and that probabilities are more similar to their 100-point rating scale, producing their observed results. They also suggest that these reversals are mediated by anchoring and insufficient adjustment. Specifically, they argue that the more compatible element serves to generate an anchor on the response scale, and that insufficient adjustment contributes to the observed increased weight for the compatible attribute.

In this paper we concentrate on a third type of compatibility effect, semantic compatibility, which involves inputs and response modes that are meaningfully related but are not expressed on the same continuum. According to Tversky *et al.* (1988), "The rationale for this principle is that characteristics of the task and the response scale prime the most compatible features of the stimulus. (p. 376)." Tversky *et al.* (1988) give as an example the discrepancy between similarity and dissimilarity judgments (Tversky, 1977). Features that two items hold in common loom larger in similarity judgments, while distinctive features loom larger in dissimilarity judgments because common features are more semantically related to the concept of similarity than are unique features.

Although demonstrations of preference reversals that can be explained in terms of strategy and scale compatibility are plentiful (Fischer & Hawkins, 1993; Goldstein & Einhorn, 1987; Grether & Plott, 1979; Hawkins, 1994; Johnson, Payne, & Bettman, 1988; Lichtenstein & Slovic, 1971, 1973; Schkade & Johnson, 1989; Tversky, Slovic, & Kahneman, 1990), there are fewer demonstrations of the semantic compatibility principle. Note that semantic compatibility relies on neither the differences in the kind of strategy used to express preferences or on the surface similarity of the response scale to one of the features of the stimulus to be judged. Rather, the semantic compatibility principle relies on the priming by the response scale of features of the object being judged. The purpose of the present experiments was to explore a new type of preference reversal that can be explained by semantic compatibility.

In two experiments subjects evaluated two types of items: consumer commodities (e.g., 1 day of vacation in Bermuda) and health-related items (e.g., a new treatment that would give you perfect 20/20 vision). Evaluations were given in terms of money and in terms of life expectancy. We hypothesized that health items would be more compatible with the life expectancy response scale and that commodities would be more compatible with the monetary scale. Although health items are not expressed on the same continuum as life expectancy, we hypothesized that life expectancy would be seen as similar to, or a good substitute for, the health items. Similarly, money would be seen as similar to, or a good substitute for, commodities. Consequently, health items would be given more weight in life expectancy evaluations and commodities would be given more weight in monetary evaluations. These differences would lead to a particular type of preference reversal in which a commodity item is preferred to a health item in monetary evaluations, but the health item is preferred to the commodity item in life expectancy evaluations.

Monetary evaluations are frequently elicited in preference reversal experiments. Life expectancy evaluation is less typically used in preference experiments, but serves as a useful second response mode in that, like money, it is a cardinal scale familiar to our subjects. Utility for life expectancy is very likely monotonic for the range of answers given by our subjects (up to 5 extra years of life expectancy). In addition, everyday decisions often involve life expectancy trade-offs, although perhaps not explicitly. For example, a decision to smoke shortens life expectancy, while a decision to eat a low-fat diet lengthens life expectancy. Unlike a certain monetary amount, however, an amount of additional life expectancy represents the mean of a dis-