Reasoning about outcome probabilities and values in preference reversals

Selart, Marcus; Boe, Ole; Gärling, Tommy

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

THINKING AND REASONING, 1999, *5* (2), 175–188

Reasoning About Outcome Probabilities and Values in Preference Reversals

Marcus Selart, Ole Boe, and Tommy Gärling
Göteborg University, Sweden

Research on preference reversals has demonstrated a disproportionate influence of outcome probability on choices between monetary gambles. The aim was to investigate the hypothesis that this is a prominence effect originally demonstrated for riskless choice. Another aim was to test the structure compatibility hypothesis as an explanation of the effect. The hypothesis implies that probability should be the prominent attribute when compared with value attributes both in a choice and a preference–rating procedure. In Experiment 1, two groups of undergraduates were presented with medical treatments described by two value attributes (effectiveness and pain-relief). All participants performed both a matching task and made preference ratings. In the latter task, outcome probabilities were added to the descriptions of the medical treatments for one of the groups. In line with the hypothesis, this reduced the prominence effect on the preference ratings observed for effectiveness. In Experiment 2, a matching task was used to demonstrate that probability was considered more important by a group of participating undergraduates than the value attributes. Furthermore, in both choices and preference ratings the expected prominence effect was found for probability.

INTRODUCTION

Research on the *preference reversal* phenomenon has contributed to an understanding of attribute weighting in judgement and choice (Lichtenstein & Slovic, 1971, 1973; Lindman, 1971; see Slovic & Lichtenstein, 1983, for review). A preference reversal occurs when an individual prefers one risky option in one procedure but reveals another preference order in another procedure. For instance, when presented with two risky options with similar expected values, participants tend to choose an option with high probability of winning a modest amount of money over an option with low probability of winning a large amount. However, participants ask for a higher minimum selling

Requests for reprints should be sent to Dr Marcus Selart, Göteborg University, Department of Psychology, PO Box 500, SE-405 30 Göteborg, Sweden.

A preliminary version of this paper was presented at the 14th research conference on subjective probability, utility, and decision making (SPUDM14), Aix-en-Provence, France, August 22–27 1993.

This study was supported financially by grants to the first author from the Swedish Council for Research in the Humanities and Social Sciences.

price for the latter. Thus, different results are obtained whether participants make a choice between the options presented to them or price each option separately.

As proposed by Slovic, Griffin, and Tversky (1990), one cause of preference reversals is the scale compatibility between attribute (outcome probability or value) and response (see Shafir, 1995, for review). As monetary value is compatible with the pricing response, the weight of this attribute is enhanced by the response mode. This explains why monetary value is more important than probability when participants indicate a selling price. However, scale compatibility does not explain why probability is more important than monetary value in choices (Schkade & Johnson, 1989). Therefore, Slovic et al. (1990) assumed that probability is a prominent attribute which receives a disproportionate weight when choices are made. The phenomenon is referred to as the *prominence effect.* This effect has previously been established for both choices and preference ratings of options with non-probabilistic outcomes (Fischer & Hawkins, 1993; Selart, 1996, 1997; Selart, Gärling, & Montgomery, 1998; Selart, Montgomery, Romanus, & Gärling, 1994; Slovic, et al., 1990; Tversky, Sattath & Slovic, 1988). In risky contexts, probabilities may be seen as a competing attribute to outcome value.

An important goal of the present research was to investigate the structure compatibility hypothesis as a complementary framework for explaining the preference reversal phenomenon. In this framework, focus is on the compatibility between the structure of the input information and the required output. Hence, it specifies that the prominence effect of a value attribute in a riskless context is not restricted to choice procedures. According to the framework, it may also occur in judgements such as preference ratings. This is because preference ratings impose a similar information structure as do choices. In a riskless context, structure compatibility is assumed to be optimal when the complexity in input and required output is at the same level (the same number of elements and relations between elements). This implies that the complexity of the input and the output is at the same level for a matching task in which a participant has to specify what change in one attribute would compensate a given change in the other attribute. If there exists a priori an ordering of attributes by prominence, a prominent, or primary, attribute will receive more weight than a secondary attribute. However, in choices and preference ratings the required output is less complex than the input because it only consists of a single element (see Selart et al., 1998, for a discussion). When the compatibility is suboptimal in this way, the weighting of the attributes will be biased so that the prominent attribute will receive more relative weight than in the matching procedure. This is the prominence effect, implying that trade-offs made implicit by suboptimal compatibility (choice, preference ratings) reflect higher relative weighting of the more important attribute than do trade-offs elicited through conditions of optimal compatibility (matching).

Structure compatibility and preference reversals

The aim of the present research was to replicate and extend the prominence effect observed for probability. In Experiment 1, the logic of the test was as follows. First, a prominence effect was established for preference ratings of options with two riskless attributes (outcome values). This involved showing that in a matching task one attribute is more important (prominent) than the other and that accordingly it should receive a disproportionate weight in the preference ratings. Second, outcome probabilities were added to the riskless options. If probability gives rise to a prominence effect, then it follows that the observed effect should be eliminated or at least reduced. This implies that the weight placed on the prominent attribute should be less. In Experiment 2, another test was devised in which a prominence effect could be demonstrated for both choices and preference ratings. An important aim of the test was to reveal that the preference reversal phenomenon is independent of whether choices or preference ratings are used as response modes. The attributes consisted of a probability and a value attribute. By means of a matching task, it was first determined whether or not probability was a prominent attribute. Then, the existence of a prominence effect was investigated in a procedure where choices and preference ratings were made of pairs of options that had been matched to be equally attractive.

EXPERIMENT 1

In Experiment 1, relative weights placed on two attributes of outcome value were determined in a matching task employing pairs of options. In this task all the outcomes presented to the participants were certain. The weights obtained from this task were then compared to those obtained in a second task consisting of preference ratings of single options. In this second task, half of the participants were assigned to a non-probabilistic group in which the outcomes were certain, whereas the remaining participants were assigned to a probabilistic group in which outcome probabilities equal to 0.25 were presented. Demonstrating a prominence effect, participants in the non-probabilistic group were expected to place more weight on the prominent attribute than on the non-prominent attribute as compared to the matching task. In contrast, if probability is a prominent attribute (Slovic et al., 1990), a smaller or no difference in weights was expected in the probabilistic group for the attributes of outcome value.

Method

Participants. A total of 32 undergraduates at Göteborg University (16 men and 16 women) participated in return for payment. An equal number of each sex was randomly assigned to a probabilistic and a non-probabilistic group which each consisted of 16 participants.

Procedure. Participants served in groups of approximately four at a time. In one part of the experiment they performed a matching task and in another part they performed preference ratings. The order between the tasks was counterbalanced across participants. A session lasted for about 30 minutes.

In both tasks participants were asked to imagine that they would receive medical treatment for a moderately serious disease. They were told that the treatment had a certain degree of effectiveness (degree of full recovery) which was expressed on a scale ranging from 0 (no effectiveness) to 100 (full effectiveness). This attribute was assumed to be prominent. The treatment also varied in degree of pain-relief expressed on a scale ranging from 0 (no pain-relief) to 100 (complete pain-relief). This attribute was assumed to be non-prominent.

In the matching task, pairs of treatment options were presented in a booklet. In each case, one of the attribute values was missing. The participants' task was to provide the missing value so that the options appeared equally attractive. In four different subgroups with an equal number of randomly assigned participants, either the effectiveness value (assumed to be the prominent attribute) or the pain-relief value (assumed to be the non-prominent attribute) was missing for the prominent option (high effectiveness value) or for the non-prominent option (high pain-relief value). Participants were informed that the value they provided had to be equal to or higher (lower) than the value for the other option on the same attribute. This information was repeated on each page in the booklet. The treatment pairs were constructed by systematically varying the difference between the highest and lowest value on effectiveness and pain-relief in steps of 5, 10, 15, and 20. For each difference, four different pairs were prepared with the highest value on each attribute varying in steps of 5 from 35 to 50, from 40 to 55, from 45 to 60, and from 50 to 65, respectively. After having constructed the options in this way, the values that were designated as missing were deleted. The order of the attributes and the options were arranged in such a way that the missing value always appeared in the lower right cell of a matrix. For instance:

	Pain relief	Effectiveness
Treatment A	35	30
Treatment B	30	?

The preference-rating task was introduced in another booklet in which single options were presented according to individually randomised orders. For each option participants indicated on a scale ranging from 0 to 100 how much they would prefer to receive the treatment if they had a choice. Different options were constructed by factorial variation of effectiveness and pain-relief, respectively, in the steps 30, 40, 50, 60, and 70. Participants assigned to the non-probabilistic group were not given any other information, whereas those assigned to the probabilistic group were told that only in 25% of all treated cases was the treatment known to have the indicated degree of effectiveness and pain-relief. In the remaining cases no effectiveness and pain-relief, respectively, were expected. Furthermore it was pointed out that as a consequence in a certain number of cases the treatment may imply an absence of both effectiveness and pain-relief. Participants were reminded on each page that the indicated degree of effectiveness and pain-relief, respectively, would only be true in 25% of all cases.

Results

The purpose of the matching procedure was to determine the relative weights of the prominent and non-prominent attributes so that they could be compared to the weights derived from the preference rating task. In order to determine these weights it is assumed that

$$u_{P,p} + u_{P,np} = u_{NP,p} + u_{NP,np} \quad (1).$$

where $u_{P,p}$ and $u_{P,np}$ denote the attractiveness of the levels of the prominent and non-prominent attributes for the option that includes the highest value on the prominent attribute and the lowest value on the non-prominent attribute, and $u_{NP,p}$ and $u_{NP,np}$ the corresponding attractiveness values for the option that includes the lowest value on the prominent attribute and the highest value on the non-prominent attribute. If the objective attribute levels are denoted x and it is assumed that $x_{i,j} = w_j u_{i,j}$ with w_j denoting the attribute weights, then by substitution in Equation 1:

$$w_p / w_{np} = (x_{NP,np} - x_{P,np}) / (x_{P,p} - x_{NP,p}) \quad (2).$$

The weight ratios for participants' individual matching values were determined based on Equation 2. In these computations, 2.0% of all observations had to be excluded because participants provided a matching value which resulted in a difference that was equal to zero. As can be seen in Table 1, all mean ratios are larger than 1.0, confirming that effectiveness was the prominent attribute. However, the mean ratios vary depending on matching condition. A 2 (non-probabilistic/probabilistic group) × 4 (matching condition) × 5 (difference between attribute levels) mixed analysis of variance (ANOVA) with repeated measures on the last factor was performed on the log-transformed ratios. It yielded a significant main effect of matching condition, $F(3, 24) = 5.74$, $P < .01$, which also interacted reliably with difference between attribute levels, $F(12, 96) = 1.95$, $P < .05$. Ignoring these significant effects, an estimate of the weight ratios was obtained for each participant by computing the mean across all the differences between attribute levels.

TABLE 1
Mean Weight Ratios in Different Conditions of the Matching Task (Experiment 1)

Prominent Option		*Non-prominent Option*	
Effectiveness matching	*Pain-relief matching*	*Effectiveness matching*	*Pain-relief matching*
1.49	1.35	3.47	1.86

The mean preference ratings are plotted in Fig. 1. A 2 (non-probabilistic/probabilistic group) × 5 (levels of prominent attribute) × 5 (levels of non-prominent attribute) mixed ANOVA with repeated measures on the last two factors yielded significant main effects of the levels of prominent and non-prominent attributes, $F(4, 120) = 108.92$, $P < .001$, and 17.84, $P < .001$, respectively. A weak although significant interaction between the levels of the attributes was also obtained, $F(16, 480) = 2.02$, $P < .05$, possibly due to floor and ceiling effects. Furthermore, it was found that non-probabilistic/probabilistic group interacted reliably with the levels of the prominent attribute, $F(4, 120) = 7.32$, $P < .001$, whereas the interaction with the levels of the non-prominent attribute did not quite reach significance, $F(4, 120) = 1.67$, $P < .20$. The significant main effects and the interaction effects were primarily accounted for by the linear trend.

Table 2 exhibits the mean weight ratios obtained from the matching task and the preference rating task, respectively. The latter ratios were computed for individual participants by dividing the mean regression coefficient corresponding to the prominent attribute with the mean regression coefficient corresponding to the non-prominent attribute. In doing this, four participants in the non-probabilistic group and one in the probabilistic group had to be excluded because of negative or zero regression coefficients corresponding to the non-prominent attribute. Indicating a prominence effect, it may be observed that the ratios obtained from the preference ratings are higher. A 2 (non-probabilistic/probabilistic group) × 2 (matching task/preference ratings) mixed ANOVA with repeated measures on the last factor was performed on the log-transformed ratios. It yielded a significant main effect of matching task/preference ratings, $F(1, 25) = 7.45$, $P < .05$. Furthermore, substantiating that the prominence effect was stronger in the non-probabilistic than in the probabilistic group, the main effect of this factor was also significant, $F(1, 25) = 4.26$, $P < .05$. As can be seen in Table 2, this difference is primarily observed in the preference ratings. Nevertheless, the interaction between non-probabilistic/probabilistic group and matching task/preference task did not quite reach significance, $F(1, 25) < 1$.

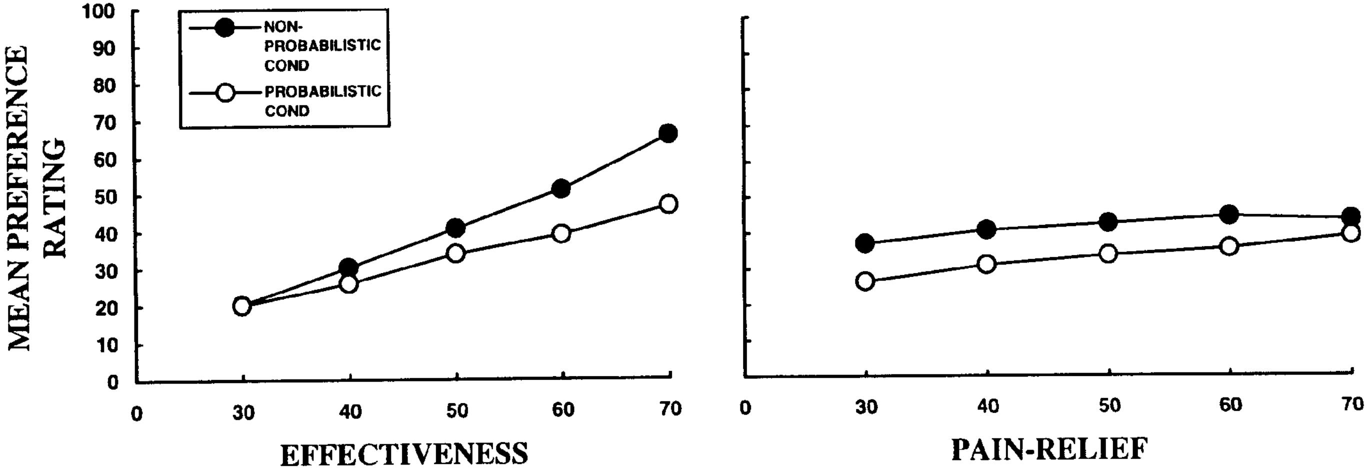

FIG. 1. Mean preference ratings in the probabilistic and non-probabilistic groups as a function of levels of the prominent and non-prominent attributes, respectively.

TABLE 2
Mean Weight Ratios in the Matching Task and Preference Ratings With and Without Outcome Probability (Experiment 1)

Matching Task[a]		*Preference Ratings*	
Outcome probability	*No outcome probability*	*Outcome probability*	*No outcome probability*
1.52	1.85	2.32	4.25

[a]In the matching task outcome probability is a dummy factor

Discussion

When comparing the preference ratings to performance in the matching task, a prominence effect was indicated by the fact that the mean ratio between the regression coefficients corresponding to the prominent (effectiveness) attribute and non-prominent (pain-relief) attribute was larger than the mean ratio observed in the matching task. Thus, participants weighted the prominent attribute relatively more in preference ratings than they did in the matching task.

In partial support of the hypothesis (Slovic et al., 1990) that probability receives disproportionate weight (thus causing a prominence effect), adding a probability to the riskless options affected the relative weights placed on the prominent and non-prominent attributes in preference ratings. As predicted, the prominent attribute received less weight in the probabilistic than in the non-probabilistic group, whereas the reverse held true for the non-prominent attribute. However, the prominence effect was not absent in the probabilistic group. Thus, although a reduction of the difference between the value attribute weights was accomplished by the introduction of probabilities, the difference between them was not entirely eliminated.

In summary, with some qualifications, the findings of Experiment 1 render credibility to Slovic et al.'s (1990) explanation of why choices are more influenced by probability than by value in preference reversals. The results also point to the fact that this phenomenon is not restricted to choices, but also applies to other response modes, such as preference ratings. As the prominence of probability is relative to another (value) attribute, an interesting question is whether it matters which this other value attribute is. In the present experiment, the prominent attribute in the non-probabilistic group considered medical treatment effectiveness. In relation to other types of attributes that are usually employed in preference reversal experiments (e.g. a small monetary amount), this attribute may be perceived as more important by the participants. This perceived importance of the value attribute may explain why a prominence effect was also observed in the probabilistic group. An implication of this is that the frequency of preference reversals would depend on how important the value attribute is.

EXPERIMENT 2

In Experiment 2, another test was devised of the hypothesis that probability is a prominent attribute if compared with a value attribute in a choice or preference-rating task. Following the same procedure as Slovic (1975) and Selart et al. (1998), participants were first asked to perform the matching task with the aim of rendering pairs of options equally attractive. Immediately thereafter, they made preference ratings and choices of the same options. Probability was always one of two attributes used to describe the options. It was expected that the matching task would reveal that probability is a prominent attribute when compared with a value attribute. Furthermore, it was expected that the choices and preference ratings would reveal a prominence effect in which probability looms larger than value.

A second aim of Experiment 2 was to investigate whether the strength of the prominence effect depends on the nature of the value attribute. If it does, the prominence effect was expected to be weaker when the value attribute consisted of treatment effectiveness than when it consisted of pain-relief. The basis for this prediction is that the former value attribute was shown in Experiment 1 to be the more important one.

Method

Participants. Another 16 male and 16 female undergraduates at Göteborg University participated in return for payment. An equal number of men and women were randomly assigned to an effectiveness and a pain-relief group, respectively, which each consisted of 16 participants.

Procedure. Participants served either individually or in small groups of at most six at a time. Upon arrival at the laboratory they were seated separately in private booths. In each booth, a PC controlled the presentation of the material. All participants were asked to imagine that they would receive medical treatment for a modestly serious disease. Their task was to evaluate the attractiveness of the treatment differing on two attributes. In one of the two groups the treatment varied in effectiveness, whereas in the other group the treatment varied in degree of pain-relief. In both groups the second attribute concerned the probability that the treatment had the indicated degree of effectiveness or pain-relief. All attributes were expressed on a scale ranging from 0 to 100. The procedure was in all other respects the same for both groups of participants. A session lasted for about 20 minutes.

On each trial, participants were first presented with two optional treatments. For any option, either the probability or the value attribute was missing. The order of presentation of attributes was controlled for in the experiment. The task consisted of typing the missing value so that the two treatments were judged to be equally attractive. After having completed all eight pairs of options, the same pairs were presented once again in the same order, although this time with the typed missing value included.[1] At this stage, the task consisted of making choices of the previously matched alternatives. After having made a choice, participants also rated their preference for one of the treatments on a scale ranging from 0 to 100. All treatment pairs were finally presented once again in the same order. In addition to making a new choice, this time participants made a preference rating of the option other than they had the first time. The presentation of preference rating options followed a counterbalanced order. Thus, the presentation order of judgement options was independent of participants' previous choices.

The different pairs of treatment options are given in Table 3. In the matching task, participants were randomly assigned to four equally large groups. In two of the groups, participants were asked to provide a missing probability value (probability matching), whereas in the other two groups they were asked to

TABLE 3
Pairs of Treatment Options Presented in the Matching Task with one Attribute Value Missing (Experiment 2)

	Prominent Option		*Non-prominent Option*	
#	*Probability*$_P$	*Value*$_P$	*Probability*$_{NP}$	*Value*$_{NP}$
1	10	5	5	10
2	20	10	10	20
3	30	15	15	30
4	40	20	20	40
5	60	30	30	60
6	70	35	35	70
7	80	40	40	80
8	90	45	45	90

The following is true: (i) Probability_p-Probability_{NP} = Value_{NP}-Value_P; (ii) Probability_{NP}/Probability_P = Value_P/Value_{NP} = 0.5. Values are either effectiveness or pain-relief.

[1]The rationale for this procedure is that in order to be consistent participants may rate the options equal in attractiveness immediately following the matching task. However, if separated in time, they should not feel equally compelled to do this and may also be unable to remember their previous preference rating. In contrast, the choice task forced participants to express a preference for one of the options. As the options were matched to be equally attractive, they were expected to be chosen equally often if there is no prominence effect. A difference between the preference ratings and choices may indicate some bias introduced by the fact that participants were forced to make a choice.

provide a missing value on the value attribute (value matching). The missing probability or value was either lower or higher than the given one on the same attribute in different groups.

Results

The mean weight ratios from the matching task were computed according to Equation 2 with probability designated as the prominent attribute. As the mean weight ratio (1.16) was larger than 1.0, the results indicated that probability was a prominent attribute (see Table 4). Furthermore, the mean ratio for effectiveness was lower (higher weight relative to probability) than the ratio for pain-relief (0.96 vs. 1.36). In a 2 (effectiveness/pain-relief group) × 2 (probability/value matching) mixed ANOVA with repeated measures on the last factor performed on the log-transformed ratios, this difference proved to be reliable, $F(1, 28) = 5.05$, $P < .05$. Furthermore, the ANOVA showed that there was a significant difference in the mean ratios between probability matchings and value matchings (1.40 vs. 0.91), $F(1, 28) = 6.68$, $P < .05$.

For choices and preference ratings a score of 1 was assigned if the prominent option (the option with the highest probability) in a pair was chosen or given the highest preference rating. If both options received the same preference ratings, a score of 0.5 was assigned. A prominence effect was observed, as the mean response scores were reliably larger than 0.50 ($P < .05$) in both groups for both methods (Table 5). A 2 (effectiveness/pain-relief group) × 2 (choice/preference

TABLE 4
Mean Weight Ratios in Different Conditions of the Matching Task (Experiment 2)

Effectiveness		*Pain-relief*	
Probability matching	*Value matching*	*Probability matching*	*Value matching*
1.21	0.71	1.60	1.12

TABLE 5
Mean Response Scores for the Different Response Tasks and Value Attributes (Experiment 2)

	Effectiveness	*Pain-relief*
Choice	0.68	0.61
Preference ratings	0.62	0.63

ratings) mixed ANOVA with repeated measures on the last factor revealed that the prominence effect did not differ between choices and preference ratings. Neither was there any reliable difference depending on whether the value attribute was effectiveness or pain-relief.

Discussion

The prediction that probability is a prominent attribute was verified by the fact that participants in the matching task placed a greater weight on probability than on the value attribute. This weight tended as expected to be higher when the value attribute consisted of pain-relief than when it consisted of effectiveness.

It was also found that in the choice and preference-rating tasks, participants more frequently preferred the option with a higher probability, although the preceding matching task had ensured that the options were equally attractive. Thus, the prediction was also verified that a prominence effect is observed in the choice and preference tasks when probability is a prominent attribute. It should be noted that this effect did not differ reliably in strength depending on whether the preference task consisted of choices or preference ratings, thereby replicating previous findings (Fischer & Hawkins, 1993; Selart, 1996, 1997; Selart et al., 1994, 1998).

Finally, no effect of the value attribute was observed on the strength of the prominence effect, despite the fact that the matching task verified the difference in weights placed on the value attributes relative to probability. In contrast to the results of Experiment 1, it thus appears as if the prominence effect does not vary in strength depending on the relative weight placed on the non-prominent attribute.

GENERAL DISCUSSION

A major aim of the present experiments was to investigate the hypothesis offered by Slovic et al. (1990) that probability is a prominent attribute and that a prominence effect would be obtained in choices between pairs of’ risky options. The results of the two experiments confirmed the hypothesis, despite the fact that the character of the present choice/preference rating task (medical treatments) differed from the ones that are usually applied in preference reversals experiments (monetary gambles). It is argued that this difference rendered the test more conservative, as it seems likely that the present value attributes (effectiveness or pain-relief of medical treatments) would be considered more important as attributes than a modest sum of money that may be won or lost in a bet. However, the prominence effect did not vary much in strength with the type of value attribute, although such a difference was observed in the matching task.

Although confirming the hypothesis offered by Slovic et al., it is argued that the preference reversal phenomenon cannot be accounted for by scale compatibility alone. This form of compatibility between attribute scale and response scale may explain why the monetary value attribute is overweighted in a pricing condition of the preference reversal task (Slovic et al., 1990). However, it fails to explain why participants act in a more "choice-like" than "pricing-like" way when they are making their preference ratings, and thus attribute the highest weight to the probability attribute.

The structure-compatibility hypothesis may therefore serve as a better explanation of the effect (Selart, 1996, 1997; Selart et al., 1998). It claims that similarities in structure compatibility between choices and preference ratings lead to the same kind of reasoning. Applying the structure-compatibility model to how participants reason in choice and pricing tasks, it is predicted that the preference reversals between choice and pricing would remain if the choices are replaced by preference ratings.

The results also have bearing on the discussion of whether or not the matching task is to be considered a completely unbiased response procedure. It has been tacitly assumed that the matching task is a preference measurement procedure that is not subject to biases. However, the present results suggest that this may not be the case. Large reliable differences in the weight ratios were in fact observed, depending on which attribute was matched. Consistent with previous research (Slovic et al., 1990), the prominent attribute in both experiments received a higher weight when it was matched. The same phenomenon was observed for the non-prominent attribute.

In conclusion, the results from the present study suggest that the reasoning about outcome probabilities and values in preference reversals is governed by a contingent-weighting mechanism (Slovic et al., 1990; Tversky et al., 1988). For this mechanism, the information structure of the problem is salient. This implies that the prominence effect of probability on choices may be explained by structure compatibility, as a similar kind of reasoning is also found when participants are making preference ratings. According to this explanation, similarities in compatibility between task requirements in input and output of a task seem to be able to explain the prominence effect. It is therefore argued that structure compatibility has explanatory value for the reasoning manifested in preference reversals.

Manuscript received 15 May 1998
Revised manuscript received 25 February 1999

REFERENCES

Fischer, G.W., & Hawkins, S.A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 580–597.

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.

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

Selart, M. (1996). Structure compatibility and restructuring in judgment and choice. *Organizational Behavior and Human Decision Processes*, *65*, 106–116.

Selart, M. (1997). Aspects of compatibility and the construction of preference. In R. Ranyard., R. Crozier., & O. Svenson (Eds.), *Decision making: Cognitive models and explanations* (pp.58–71). London: Routledge.

Selart, M., Gärling, T., & Montgomery, H. (1998). Compatibility and the use of information processing strategies. *Journal of Behavioral Decision Making*, *11*, 59–72.

Selart, M., Montgomery, H., Romanus, J., & Gärling, T. (1994). Violations of procedural invariance in judgment and choice. *European Journal of Cognitive Psychology*, *6*, 417–436.

Shafir, E. (1995). Compatibility in cognition and decision. In J. Busemeyer., R. Hastie, & D.L. Medin (Eds.), *Decision making from a cognitive perspective* (pp.247–274). San Diego, CA: Academic Press.

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 280–287.

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. In R.M. Hogarth (Ed.), *Insights in decision making* (pp.5–27). Chicago, IL: University of Chicago Press.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review*, *73*, 623–638.

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