

Structure Compatibility and Restructuring in Judgment and Choice

MARCUS SELART

Göteborg University, Sweden

The use of different response modes has been found to influence how subjects evaluate pairs of alternatives described by two attributes. It has been suggested that judgments and choices evoke different kinds of cognitive processes, leading to an overweighing of the prominent attribute in choice (Tversky, Sattath, & Slovic, 1988; Fischer & Hawkins, 1993). Four experiments were conducted to compare alternative cognitive explanations of this so-called prominence effect in judgment and choice. The explanations investigated were the structure compatibility hypothesis and the restructuring hypothesis. According to the structure compatibility hypothesis, it was assumed that the prominence effect is due to a lack of compatibility between the required output from subjects and the structure of information in input. The restructuring hypothesis stated that the decision maker uses mental restructuring operations on a representation of decision options to make the options more clearly differentiated. In Experiment 1, a matching procedure was used to provide pairs of equally attractive options (medical treatments) for the following experiments. In Experiments 2, 3, and 4, preferences were elicited with two different response modes, choice and preference rating. Value ranges on the prominent and nonprominent attributes were manipulated to test the structure compatibility hypothesis. Accountability was also subject to manipulation as it was assumed to stimulate restructuring. Since the prominence effect was not restricted to choices, and effects of value ranges were obtained but not of accountability, the results were interpreted in line with the structure compatibility hypothesis. © 1996 Academic Press, Inc.

The study was financially supported by grants to the author from the Swedish Council for Research in the Humanities and Social Sciences and from The Royal Swedish Academy of Sciences. A previous version of the paper was presented at a symposium organized jointly by the European Group for Process Tracing Studies of Decision Making and The Czech Republic Academy of Sciences, Práha, The Czech Republic, April 23–27, 1993. I thank Carl Martin Allwood, Tommy Gärling, Henry Montgomery, and Ola Svenson for valuable comments. I am also grateful to Daniel Eek, Robert Gillholm, and Joakim Romanus for assistance in collecting the data. Address correspondence and reprint requests to Marcus Selart, Department of Psychology, Göteborg University, Haraldsgatan 1, S-413 14 Göteborg, Sweden.

The principle of procedure invariance is one of the fundamental assumptions of normative decision theory. It implies that equivalent ways of eliciting decision makers' preferences should result in similar outcomes (Kahneman & Tversky, 1984). However, procedure invariance is often violated by preference reversals (e.g., Slovic, Griffin, & Tversky, 1990; Slovic & Lichtenstein, 1983, Tversky, Sattath, & Slovic, 1988; see Payne, Bettman, & Johnson, 1992, for a review). A preference reversal occurs whenever an individual prefers one alternative in one procedure but shows the opposite preference order in another. For example, when subjects have to indicate which one of approximately equally attractive bets they prefer and how much they would be willing to sell the bets for, they often choose the high-probability option but indicate the highest selling prize for the high-payoff option (Slovic & Lichtenstein, 1983). Choosing between a pair of gambles thus seems to involve different psychological processes than bidding for each one separately. When people are asked to choose between two bets, they pay particular attention to the probability of winning. When they are asked to set a price for how valuable the bet is, they look at how large the potential payoffs are.

It has been shown in a riskless context that multi-attribute options invoke different attention processes in judgment and choice (Billings & Scherer, 1988; Lindberg, Gärling, & Montgomery, 1989; Westenberg & Koele, 1990, 1992). Specifically, response-mode biases have been found when subjects are asked to evaluate pairs of decision alternatives whose consequences are described by two attributes. Recently, Slovic *et al.* (1990) and Tversky *et al.* (1988) demonstrated a judgment-choice discrepancy, or riskless preference reversal, in such a case. One of the attributes was selected to be predominant or prominent. In choice tasks subjects placed more weight on this attribute than they did in a matching task in which they were required to make the two options equally attractive. Subsequently, this "prominence" effect has been replicated by Montgomery, Gärling, Lindberg, and Selart (1990); Montgomery, Selart, Gärling, and Lindberg (1994); and Selart, Montgomery, Romanus, and Gärling (1994).

The compatibility hypothesis has been proposed to account for the prominence effect (Tversky *et al.*, 1988; Slovic *et al.*, 1990; Fischer & Hawkins, 1993; Hawkins, 1994). According to this hypothesis, the effect reflects a general principle of compatibility according to which the processing of input (e.g., attributes describing options in a judgment or choice task) depends on how compatible it is with the output (i.e., subjects' responses). Tversky *et al.* (1988) argued that both risky and riskless preference reversals are governed by stimulus–response compatibility. Identical components on both the stimulus and the response side enhance compatibility. Such components are the use of the same scale units (e.g., grades, ranks), the direction of relations (e.g., whether the correlations between input and output variables are positive or negative), and the numerical correspondence (e.g., similarity between the input and output). The use of similar scale units has been referred to as scale compatibility by Fischer and Hawkins (1993) and Hawkins (1994). The framework has also been used by Chapman and Johnson (1994), who state that scale compatibility occurs if an anchor and a preference judgment are expressed on the same scale. According to Slovic *et al.* (1990), the theoretical motivations for scale compatibility are somewhat loose but its implications seem unambiguous.

However, scale compatibility cannot alone account for the prominence effect (Hershey & Schoemaker, 1985; Schkade & Johnson, 1989; Slovic *et al.*, 1990). It may account for variations in strength of preference found in different judgment modes, but in judgment–choice comparisons this type of compatibility can yield predictions directly opposed to the prominence effect (Fischer & Hawkins, 1993). For instance, if dollar is the prominent attribute, the scale-compatibility hypothesis implies that people will attach greater weight to money in a dollar-matching task than in choice.

In line with this criticism, Chapman and Johnson (1995) state that it is not enough to consider only the surface similarity of the response scale to one of the features of the stimulus. The reason for this is that the semantic categorizations of the involved objects (e.g., money vs health improvement) also must be considered (see also Tversky, 1977).

Another recent explanation of the prominence effect can also be related to the results of Tversky *et al.* (1988). It states that the effect occurs because choice and matching tasks evoke different types of decision strategies which give different weight to the prominent attribute. The qualitative response in choice is regarded as compatible with a lexicographic decision rule which renders quantitative weighting of attributes unnecessary. In contrast, quantitative judgments are compatible with a quantitative weighting rule. This form of compatibility is referred to as strategy compat-

ibility (Fischer & Hawkins, 1993; Hawkins 1994). The cognitive basis of this is that the response mode is priming the focus of attention on the compatible features of the input. Nevertheless, the idea of strategy compatibility has also proved to have weaknesses. For instance, it has been shown that both choices as well as several kinds of judgments evoke a prominence effect (Fischer & Hawkins, 1993; Selart *et al.*, 1994)

A different perspective on compatibility takes as its point of departure the information structure of judgments and decisions (Montgomery *et al.*, 1994; Selart *et al.*, 1994, Selart, 1994a, 1994b). From this point of view, it is assumed that the prominence effect is due to a lack of compatibility between the required output from subjects and the structure of information in input. The general idea is that matching judgments can be seen as distinct from preference rating judgments and choices in that subjects have to evaluate one value difference relative to another to carry out the task. In matching judgments, the task itself prevents the decision maker from using a lexicographic strategy, which is not the case in choice and preference rating. This depends on that subjects in the matching task have to fill in a missing value in the matrix of inputs, whereas the same value is given as an input in choice and preference rating. Thus, in matching judgments there is a compatibility between input and output which is lacking in choice and preference rating. On the basis of this reasoning, a prominence effect is also expected for preference rating, due to the assumption that information structure compatibility is salient for the selection of strategy. This explanation is henceforth referred to as structure compatibility.

The suggestion has also been made that the prominence effect can be explained by a hypothesis based on Montgomery's (1983) theory of dominance structuring in decision making. This hypothesis states that the decision maker uses mental restructuring operations on a representation of decision options to make the options more clearly differentiated. These restructuring operations are parts of a fundamental cognitive process in which humans apply operations to information to yield a new problem representation (Payne, Bettman, Coupey, & Johnson, 1992). In this theory, it is assumed that subjects making choices restructure the available information to make one option dominate the other(s). Subjects may therefore increase value differences between options on important attributes and decrease differences on unimportant attributes.

Similar assumptions have more recently been made by Svenson (1992). In his differentiation and consolidation theory, decision making is modeled as a process in which one alternative gradually is differentiated from another until the degree of differentiation is sufficient for a decision. Thus, it is not sufficient to choose

an alternative that is better. Rather, the preferred alternative should be so much better than the nonpreferred alternative that it remains better even under unfavorable post-decision conditions.

If there are one prominent and one nonprominent attribute, as a result of the restructuring process, the former will have more influence since the differences on that attribute are enlarged relative to the other. More precisely, subjects may modify their beliefs or values in such a way that there will be a larger discrepancy between the options on the more important attribute than on the less important attribute. Both the importance order of the attributes and the differences between the alternatives on the attributes will then speak in favor of a preference in line with the prominent attribute. The above-mentioned modifications of beliefs and values are to be particularly frequent in choice which is characterized by a conflict between alternatives. It is assumed that reasons or motives guide the modifications of values and beliefs. Therefore, both Montgomery (1983) and Svenson (1992) assume that the importance of a decision is directly related to the degree of restructuring.

EXPERIMENTAL HYPOTHESES

A prominence effect is predicted in both choice and preference ratings according to the structure compatibility hypothesis. The hypothesis can also be tested by varying value differences between the options. Narrow value ranges are supposed to decrease the effect in both choices and preference ratings, whereas wide ranges should increase it. This is because an information structure characterized by narrow value ranges between options may be assumed to invoke evaluations of one attribute difference in relation to another as in matching. Narrow value ranges are supposed to prime a similar compensatory strategy, leading to a reduction of the prominence effect. Wide value ranges should, on the other hand, prime a noncompensatory strategy, leading to the evaluation of primarily the difference on the prominent attribute.

According to the restructuring hypothesis, a prominence effect is predicted to occur in choice, as was assumed in the former hypothesis. To further test the restructuring hypothesis, it is proposed that accountability will increase restructuring in the choice condition (Simonson, 1989; Simonson & Nye, 1992; Tetlock, 1992). This is because justifying reasons for a choice should influence modifications of values and beliefs. Hence, a larger prominence effect is predicted for choices of high accountability than for those of low accountability. In line with this, other research suggests that people cope with accountability by seeking out the most acceptable position, which in a situation like this would be synonymous with giving a higher weight to

the prominent attribute (Pruitt, 1981; Cialdini, Levy, Herman, & Evenbeck, 1973). According to the structure compatibility hypothesis, the manipulation is however not supposed to affect the prominence effect.

In Experiment 1, a matching procedure was used to provide the stimulus material (medical treatments) for the following experiments. In Experiments 2, 3, and 4, preferences for the medical treatments were elicited with two different response modes, choice and preference rating. Two partly different manipulations of accountability were used in these experiments. Value range was varied as a between-subjects factor in Experiment 2 and as a within-subject factor in Experiments 3 and 4.

The choice of making between-group comparisons of the initial matching task with the latter choice/preference rating experiments was based on that Tversky *et al.* (1988) used between-group comparisons of matching and choice conditions as a basis for the explanations of the prominence effect (strategy compatibility, scale compatibility).

EXPERIMENT 1

Method

Subjects. Forty undergraduate students at Göteborg University, equally as many men as women, participated in return for payment. Ten subjects were randomly assigned to each of four matching conditions so that equally as many men as women performed four different versions of the matching task.

Stimuli. The stimuli consisted of matching problems involving two medical treatments which only differed in effectiveness and pain relief. Effectiveness was assumed to be the more important or prominent attribute. Values on both attributes were expressed on a scale ranging from 0 (no effectiveness/pain relief) to 100 (full effectiveness/pain relief). Pairs of alternatives were constructed by systematically varying the range between the highest and lowest value on each attribute in steps of 5, 10, 15, 20, or 25. For each range four replicate 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, from 50 to 65, and from 55 to 70, respectively. After having constructed the options in this way, the values to be filled in were deleted as will be described below. All the items are listed in the Appendix (Table A1).

Procedure. Subjects served in groups of approximately four at a time and were instructed to carry out the task individually. All pairs of descriptions of treatments were presented to them in a booklet. Ten random orders were used equally often. On average the matching tasks were completed in 15 min.

Subjects were asked to imagine that they were suf-

fering from a disease. On the first page of the booklet, the type-written instructions read as follows: "Suppose you need medical treatment for a disease. There exist two treatment programs which only differ in effectiveness and pain-relief. The definition of effectiveness is to what degree you will be fully recovered. The degrees of effectiveness and pain-relief are expressed on a scale ranging from 0 (no effectiveness/pain-relief) to 100 (full effectiveness/pain-relief). Please evaluate the following options." On each following page of the booklet, a pair of treatments were shown. For one of the treatments one attribute value was missing. For instance,

	Pain relief	Effectiveness
Treatment A	60	45
Treatment B	45	?

Subjects' task was to provide the missing value so that the options were experienced as equally attractive. In the example subjects thus indicated how effective Treatment B must be to appear equally attractive as Treatment A. They were informed that the value provided had to be equal to or higher (lower) than the value for the other option on the same attribute.

In four different matching conditions varied as a between-subjects factor, the missing value was either the highest or the lowest on the prominent attribute or the highest or the lowest on the nonprominent attribute. The pairs of treatments were presented in a matrix with options as rows and attributes as columns. The order of the attributes and options were in the different conditions arranged in a way so that the missing value always appeared in the lower right cell of the matrix.

Results and Discussion

The analyses of the results rested on the assumption 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 nonprominent attributes for the prominent option (with the highest value on the prominent attribute), and $u_{NP,p}$ and $u_{NP,np}$ denote the corresponding attractiveness of the levels of the prominent and nonprominent attributes for the nonprominent option. If the objective attribute levels are denoted x and it is assumed that $u_{i,j} = w_j x_{i,j}$ with w_j denoting the attribute weights, then by substitution in Eq. (1):

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

Based on Eq. (2), the weight ratios were determined for individual subjects' matching values. In these computations, 6.0% of all observations were excluded because subjects provided a matching value which resulted in a range which was 0. The mean weight ratios

ranging between .83 and 3.73 for the different matching conditions and value ranges with an overall mean of 1.89 indicated that effectiveness was a more important attribute than pain-relief. A 4 (matching condition) \times 5 (value range) ANOVA with repeated measures on the last factor yielded no reliable effects of value range. The construction of the stimuli for the following experiments was accordingly based on the overall mean weight ratio.

EXPERIMENT 2

Method

A way of testing the prominence effect is to create a set of problems so that subjects can either choose or rate their preference for pairs of options which have been matched to be equally attractive (Montgomery *et al.*, 1994).

Subjects. Eighty undergraduate students at Göteborg, University, equally as many men as women, participated in return for payment. Five men and five women were randomly assigned to each of the eight conditions. Subjects were run in groups of four in an order which was counterbalanced.

Stimuli. Pairs of descriptions of medical treatments were constructed as follows. In one set of pairs the value ranges on the prominent attribute was 5 (narrow value ranges); in another set, they were 40 (wide value ranges). A value range on the nonprominent attribute was for each treatment pair obtained by first multiplying the value range on the prominent attribute with the mean weight ratio obtained from the matching task in Experiment 1. Using a similar procedure as Tversky *et al.* (1988), the resulting value range was then increased by 1 to make the nonprominent option appear more attractive than the prominent option. Twenty pairs of options were created for narrow and wide value ranges. In both conditions, the attribute levels of the prominent option on the prominent attribute varied in steps of 1 from 41 to 60. Each option was presented on a single page. Two examples of each are given in Table 1.

Procedure. Preferences were elicited with choice and preference rating. In the choice response mode, subjects made 20 choices between the pairs of treatments. The type-written instructions were the same as in the matching task except that subjects were asked to choose between the options. Subjects indicated which treatment they would choose.

The descriptions were presented pairwise in the preference-rating condition like they were in the choice condition. In this response mode, subjects rated each option in the pairs on a scale ranging from 0, defined as very poor, to 100, defined as very good. Subjects were

TABLE 1

Examples of Stimuli Used in Experiments 2 and 3

Value ranges	Options	Prominent attribute	Nonprominent attribute
Narrow	Treatment A	41 ^a	36
	Treatment B	36	46
Narrow	Treatment A	42	37
	Treatment B	37	47
Wide	Treatment A	41	1
	Treatment B	1	78
Wide	Treatment A	42	2
	Treatment B	2	79

^a The values of the prominent and the nonprominent attributes are expressed on a scale ranging from 0 (very low) to 100 (very high).

instructed that they were allowed to use any value on the scale to indicate their preference.

Accountability was manipulated through task instructions in a way similar to that in previous studies (Simonson, 1989; Simonson & Nye, 1992; Tetlock, 1983). In the low accountability condition, subjects were told that their responses would remain totally confidential. In contrast, subjects were in the high accountability condition told that the material they provided by their responses would be included in a booklet and subject to future class discussions. They were therefore asked to explain and justify their judgments or choices. After having completed all choices/ratings, subjects were allowed to review their responses and then write down their explanations and justifications. Without having been told in advance, subjects in the low-accountability condition were asked to do the same.

Subjects participated individually in sessions lasting for about 40 minutes. They were randomly assigned to one of eight conditions according to a 2 (response mode) \times 2 (accountability) \times 2 (value range) factorial design. The order of the choice problems was randomized. Across subjects the orders of prominent/nonprominent attribute and option were counterbalanced.

Results and Discussion

Choices and preference ratings were scored equivalently. A score of 1 was assigned if the prominent option in a pair was chosen or given the highest preference rating. If both options received the same preference rating, a score of 0.5 was assigned. As indicated by the fact that the mean response scores were reliably larger than .50 in all conditions, a prominence effect was uniformly obtained. The weakest effect was obtained for preference ratings of options with narrow value ranges, ($p < .01$).

A 2 (response mode) \times 2 (accountability) \times 2 (value range) ANOVA indicated that the prominence effect was reliably weaker when the value range was narrow

than when it was wide, $F(1,72) = 5.68, p < .05$ (Fig. 1). In addition the interaction with response mode was significant, $F(1,72) = 4.48, p < .05$. Tukey post hoc tests revealed that the response scores for wide value ranges differed significantly from the scores from narrow value ranges only when subjects were making preference ratings.

There were no main or interaction effects of accountability. However, the interaction between response mode and accountability was marginally significant, $F(1,72) = 2.82, p < .10$ (Fig. 2).

In order to determine whether there were differences in restructuring in the different conditions, subjects' written explanations/justifications were treated similarly as think-aloud data in previous studies (Montgomery *et al.*, 1994; Selart *et al.*, 1994). Briefly, the written explanations/justifications were first partitioned into statements which normally corresponded to a sentence. Second, each statement was coded with respect to (i) which of the attributes, if any, it referred to, and (ii) whether it was positive, negative, or neutral. A statement in which the subject compared the attractiveness of both attributes was coded as a positive evaluation of the preferred and a negative evaluation of the nonpreferred attribute. The reliability of the coding was satisfactory as indicated by an 85% agreement for a randomly chosen 10% of all statements which were coded by an additional judge.

Since the justifications and explanations took part after the judgments or choices, subjects did not refer to specified alternatives or attribute levels. Rather, they expressed their explanations/justifications in terms which could be interpreted as positive or negative evaluations of the attributes. Hence, only evaluation of

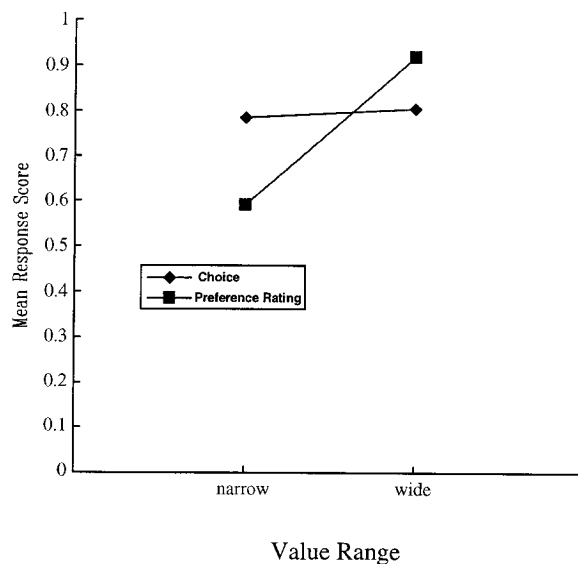


FIG. 1. Mean response scores as function of response mode and value ranges in Experiment 2.

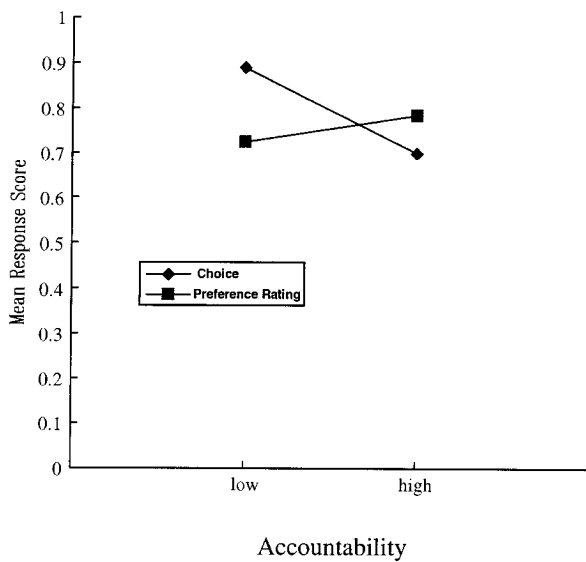


FIG. 2. Mean response scores as function of response mode and accountability in Experiment 2.

attributes were coded. The mean number of times the prominent and nonprominent attributes were positively and negatively evaluated without any reference to alternatives or attribute levels are given in Table 2. As revealed by a 2 (response mode) \times 2 (accountability) \times 2 (value range) by 2 (attribute) ANOVA with repeated measures on the last factor, the prominent attribute received a significantly higher evaluation than the nonprominent attribute, $F(1, 72) = 55.11, p < .001$. Furthermore, both attributes received higher evaluations in preference ratings than in choices $F(1, 72) = 4.39, p < .05$. This tendency was most pronounced in the case of low accountability as substantiated by an interaction between response mode and accountability, $F(1, 72) = 4.39, p < .05$.

Since evaluations of attribute levels could not be coded, the analyses do not provide a proper test of the restructuring hypothesis in this respect. However, a more abstract notion of restructuring could be tested that has been labeled restructuring by differentiation through attribute importance (Svenson, 1992). Subjects do not only restructure by changing the positions

TABLE 2
Mean Evaluations of Attributes in Experiment 2

Condition	Choice		Preference rating	
	PA ^a	NPA ^b	PA	NPA
Low accountability				
Narrow value range	1.10	-0.70	1.60	0.20
Wide value range	1.70	-0.90	2.00	-0.20
High accountability				
Narrow value range	1.50	-0.40	1.30	-0.70
Wide value range	1.30	-0.60	1.50	-0.30

^a Prominent attribute.

^b Nonprominent attribute.

of attribute levels, but they may also restructure the importance of a whole attribute.

In summary, there was no clear evidence indicating that accountability affected the prominence effect. Neither did the results indicate that accountability leads to more restructuring. It may be the case that the accountability manipulation was too weak.

EXPERIMENT 3

Method

In this experiment an attempt was made to increase the impact of the accountability manipulation by asking subjects to justify their responses each time after having made them. This form of accountability manipulation has been used in previous research (Simonson & Nye, 1992) and has in itself been able to produce a higher level of accountability even without the explicit information that the choices/ratings would be evaluated.

Subjects. Forty undergraduate students at Göteborg University, equally as many men as women, participated in return for payment. Five men and five women were randomly assigned to each of the four conditions.

Stimuli. The stimuli were a subset of those in Experiment 2. Ten pairs of treatments were taken from the set of treatments with wide value ranges and another ten from the set with narrow value ranges. Accordingly, value range was treated as a within-subject factor. The orders of prominent/nonprominent attribute and option were counterbalanced across options for each subject.

Procedure. The procedure was identical to that in Experiment 2, except that accountability was manipulated differently. As in Experiment 2, subjects in the low-accountability condition were instructed that their responses would remain totally confidential, whereas in the high-accountability condition, subjects were instructed that their choices would be included in a booklet to be used as a basis for further class discussions. In contrast to Experiment 2, subjects in the high-accountability condition were asked to explain and justify in a written statement each choice or pair of preference ratings immediately after elicitation. No explanations/justifications were required from subjects in the low-accountability condition.

Results and Discussion

The same scoring procedure was used as in Experiment 2. Since the mean response scores were reliably larger than .50 in all conditions, a prominence effect was again uniformly obtained. The weakest effect was obtained for choices between options with low accountability, ($p < .01$). A 2 (response mode) \times 2 (accountability) \times 2 (value range) ANOVA with repeated measures

on the last factor indicated, like that in Experiment 2, that the prominence effect was reliably smaller when the value range was narrow than when it was wide, $F(1, 36) = 5.00, p < .05$ (Fig. 3). However, neither the main nor the interaction effect of accountability reached significance (Fig. 4). In a second analysis all responses in the high accountability condition that were not accompanied by verbal explanation or justification were deleted. Nevertheless, this did not result in significant effects of accountability, and the main effect of value range persisted as was revealed by an ANOVA based on the remaining justified responses $F(1, 36) = 6.31, p < .05$.

The analysis of the explanations/justifications followed the same procedure as in Experiment 2. However, it had to be limited to high-accountability subjects. One of the subjects was missing because he or she only partially produced explanations and justifications. The means of positive and negative statements which referred to attributes are given in Table 3. As revealed by a 2 (response mode) \times 2 (attribute) \times 2 (value range) ANOVA with repeated measures on the last two factors, the prominent attribute received significantly higher evaluations than the nonprominent attribute, $F(1, 17) = 27.31, p < .001$.

EXPERIMENT 4

Method

In this experiment, another attempt was made to increase the impact of the accountability manipulation by instructing subjects to justify and explain their evaluations as if they were physicians. Simonson and Nye (1992) found that accountability affected the prominence effect only if selection of the normatively

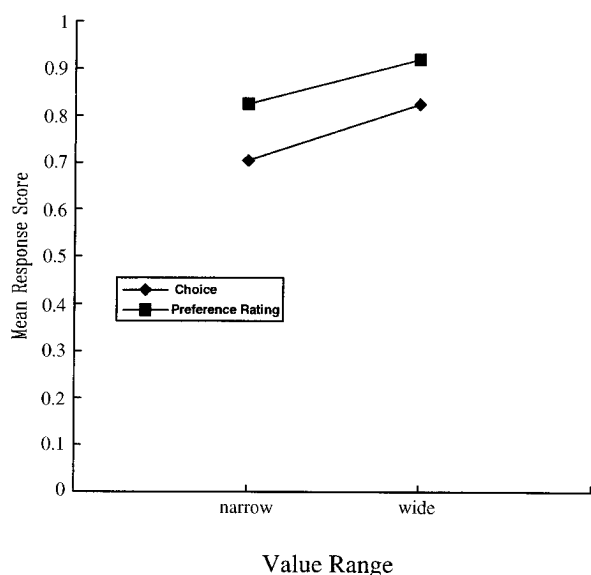


FIG. 3. Mean response scores as function of response mode and value ranges in Experiment 3.

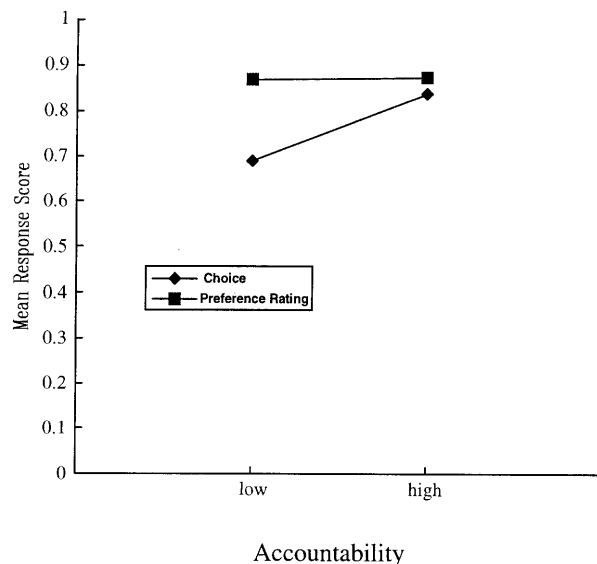


FIG. 4. Mean response scores as function of response mode and accountability in Experiment 3.

correct option was easier to justify. Previous research have suggested that accountability may be regarded as inherent in the role of clinical professionals (Greenblatt, 1991; O'Neill, 1989). By letting accountable subjects evaluate the options as physicians should therefore further increase the normatively correct appearance of the prominent option.

Subjects. Forty-eight undergraduate students at Göteborg University, equally as many men as women, participated in return for payment. Six men and six women were randomly assigned to each of the four conditions.

Stimuli. The stimuli were the same as those used in Experiments 2 and 3, with two exceptions. First, both high and low accountability subjects were instructed to take the role of the physician in making their judgments and decisions. Second, as in the former experiments, a value range on the nonprominent attribute was obtained for each treatment pair by first multiplying the value range on the prominent attribute with the mean weight ratio obtained from the matching task in Experiment 1. The resulting value range was increased by 5 to make the nonprominent option appear even more attractive (Table 4). Value ranges were treated as a within-subject factor as in Experi-

TABLE 3
Mean Evaluations of Attributes in Experiment 3

Value ranges	Choice		Preference rating	
	Prominent attribute	Nonprominent attribute	Prominent attribute	Nonprominent attribute
Narrow	2.10	-1.50	1.11	-1.11
Wide	2.70	-2.70	1.44	-1.33

ment 3. The orders of prominent/nonprominent attribute and option were counterbalanced across options for each subject.

Procedure. The procedure was identical to that in Experiment 3. Subjects in the low-accountability condition were instructed that their responses would remain totally confidential, whereas in the high-accountability condition, subjects were instructed that their choices would be included in a booklet to be used as a basis for further class discussions. Subjects in the high-accountability condition were asked to explain and justify, in a written statement, each choice or pair of preference ratings immediately after elicitation. No explanations/justifications were required from subjects in the low-accountability condition.

Results and Discussion

The same scoring procedure was used as in Experiments 2 and 3. The mean response scores were reliably larger than .50 in a majority of the conditions. The weakest effect was obtained for choices between options with high accountability ($p < .01$). A lack of a prominence effect was found for choices between options with narrow value ranges. A 2 (response mode) \times 2 (accountability) \times 2 (value range) ANOVA with repeated measures on the last factor indicated, like those in Experiments 2 and 3, that the prominence effect was reliably smaller when the value range was narrow than when it was wide, $F(1, 44) = 25.82, p < .0001$ (Fig. 5). Furthermore, a reliable interaction effect between response mode and value range was obtained, $F(1, 44) = 4.42, p < .05$. However, neither the main nor the interaction effect of accountability reached significance (Fig. 6). In a subsequent analysis, all responses in the high-accountability condition that were not accompanied by verbal explanation or justification were deleted. Nevertheless, this did not result in any significant effects of accountability, and the main effect of value range persisted as was revealed by an ANOVA

based on the remaining justified responses, $F(1, 44) = 4.86, p < .05$.

The analyses of the data obtained from high accountability subjects' explanations/justifications followed the same procedure as in Experiment 3. Table 5 shows the means of positive and negative statements which referred to the attributes. A 2 (response mode) \times 2 (attribute) \times 2 (value range) ANOVA with repeated measures on the last two factors did not yield any main effects. However, a reliable interaction between attribute and value range was obtained, $F(1, 22) = 5.12, p < .05$. The prominent attribute was more positively evaluated for options with wide value ranges, whereas the nonprominent attribute was more positively evaluated for options with narrow value ranges.

GENERAL DISCUSSION

The structure compatibility hypothesis made three predictions. First, it predicted a prominence effect in both choice and preference rating judgment. As revealed by previous research (Fischer & Hawkins, 1993; Montgomery *et al.*, 1994; Selart *et al.*, 1994), an effect was found also for preference ratings. In these studies, a large variety of problems were used. The present finding therefore supported the notion of the generality of the effect and suggests that, for instance, strategy compatibility between input and output information cannot alone account for the effect. The power of the effect was revealed in that the attempts to eliminate preferences in line with the effect proved to be unsuccessful.

The explanation for this is, according to the structure compatibility hypothesis, that the required output from subjects needs to be compatible with the structure

TABLE 4
Examples of Stimuli Used in Experiment 4

Value ranges	Options	Prominent attribute	Nonprominent attribute
Narrow	Treatment A	41 ^a	36
	Treatment B	36	50
Narrow	Treatment A	42	37
	Treatment B	37	51
Wide	Treatment A	41	1
	Treatment B	1	82
Wide	Treatment A	42	2
	Treatment B	2	83

^a The values on the prominent and nonprominent attributes are expressed on a scale ranging from 1 (very low) to 100 (very high).

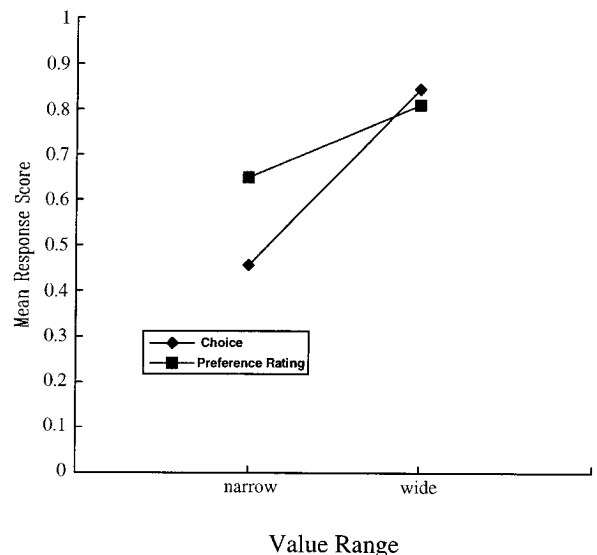


FIG. 5. Mean response scores as function of response mode and value ranges in Experiment 4.

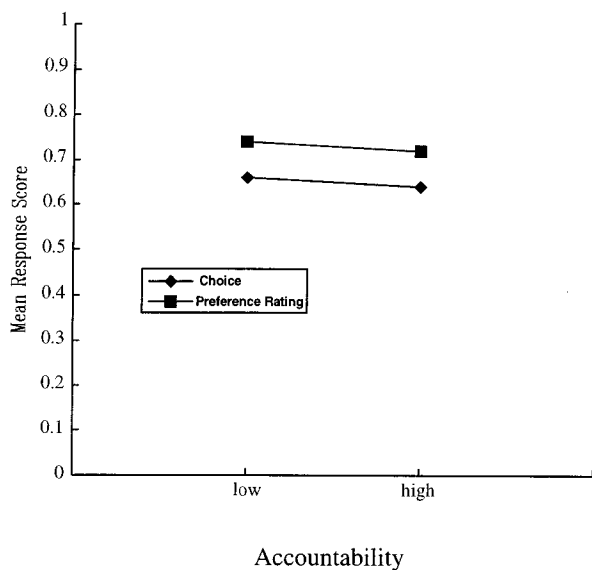


FIG. 6. Mean response scores as function of response mode and accountability in Experiment 4.

of information in input. In a matching task, there is an agreement between input and the required output. Subjects are required to match one value difference (required output) to another difference which is given in the task (input). Hence, there exists a dimensional compatibility between input and output. If the difference between attribute levels of the prominent attribute serves as input, then the difference between the levels on the nonprominent attribute serves as the required output and vice versa. This form of compatibility involves transforming and rearranging, since both differences always have to be taken into account.

In choices and preference ratings, on the other hand, both differences serve as input. For example, in a two options by two attributes choice task, the input is built on four pieces of information (the attribute levels of each alternative), whereas the output corresponds to a preference order between alternatives. Here, there is no dimensional compatibility between input and output. Subjects therefore do not transform and rearrange the information to the same extent, simply because they are not forced to do it. The information structure of these tasks gives them the opportunity to select a lexicographic strategy.

Recent findings suggest that decision strategies are

TABLE 5
Mean Evaluations of Attributes in Experiment 4

Value ranges	Choice		Preference rating	
	Prominent attribute	Nonprominent attribute	Prominent attribute	Nonprominent attribute
Narrow	-0.33	1.08	1.33	-1.25
Wide	1.50	-1.42	1.33	-0.83

influenced by the organization of individual items of information into structures (Schkade & Kleinmuntz, 1994). It is further assumed that organization as a feature of information strongly influences information acquisition. For instance, presenting the stimulus on a screen by alternative or by attribute has been shown to strongly influence information acquisition. Hence, similar task demands in both choice and preference rating (the lack of structure compatibility) can be seen as representing a common organization principle which is different from the one used in matching.

However, in Experiment 4, there was no prominence effect revealed in the choice condition for narrow value ranges. In line with this, the analyses of high-accountability subjects' explanations/justifications showed that the evaluation of the prominent attribute for this condition was slightly negative, which was not the case in the other experiments. This must nevertheless be regarded as a special case since the nonprominent option was made heavily dominant in a situation of great conflict (narrow value ranges).

Second, the structure compatibility hypothesis predicted that narrow value ranges should decrease the effect in both choice and preference ratings, whereas wide value ranges should increase it in both response modes. The rationale behind this prediction was that narrow value ranges should to a higher extent force the subjects to take all the attribute levels into account as in the matching task. As suggested by the hypothesis, reliable main effects of value range in the predicted direction were generally obtained.

Therefore, it seems plausible to assume that the increased prominence effect due to wide value ranges can be attributed to the same principle as in choice and preference rating. This form of compatibility explanation suggests that wide value ranges make subjects to a higher extent use a lexicographic strategy, that is, they overweight the prominent attribute.

An alternative interpretation of the obtained results is that the value range effect could be explained by an additive model. This explanation is based on that the utility values of the treatments which are narrow in range are more similar than the values of those treatments which are wide in range. However, two objections can be raised against this latter suggestion: (a) A prominence effect was obtained also when value ranges were narrow, despite that the utility values were more similar in this condition. (b) An additive model predicting a prominence effect due to wide ranges should also assume an overweighting of the prominent attribute in the matching task, due to these ranges. No such an effect was obtained.

Third, the structure compatibility hypothesis predicted that accountability was supposed to neither increase nor decrease the effect in choice. In line with the prediction, no reliable main effects of accountability

were found in the experiments, despite the reported tendencies.

Restructuring operations, on the other hand, imply a usage of different kinds of operations, such as bolstering the advantages or deemphasizing the disadvantages of a promising option. Thus, the restructuring hypothesis predicted a larger prominence effect for high accountability choices than for low accountability choices. However, the results revealed that the accountability manipulations did not reach significance in any of the experiments. There are several possible explanations for this. One is that the manipulations were too weak. Another builds on the results of Simonson and Nye (1993). They found that high accountability leads to an enhanced prominence effect, but that low accountability also could lead to such an enhancement if the rational alternative appeared easy to explain and justify. In Experiment 2, the prominence effect in choice was slightly increased in the low-accountability condition as compared to the high-accountability condition. A possibility is that the rational alternative in the present study was easy to explain and justify, and that subjects therefore did not restructure in the high-accountability condition. People in this situation might cope with accountability by thinking in more multidimensional ways that damage the process of restructuring. According to this interpretation, subjects treat the attributes as compensatory, that is, a high value on one attribute might compensate for a low value on another. In line with this, other research has shown that accountability may lead to an increased willingness to pay attention to information, to reduced

overattribution effects, and also to reduced overconfidence effects (Hagafors & Brehmer, 1983; Tetlock, 1983; Tetlock, 1985).

In Experiment 3, an alternative accountability instruction was used in which subjects in the high-accountability condition were instructed to justify and explain each response during the session. This manipulation was similar to that used in previous research by Simonson and Nye (1992). They found that subjects who justified their choices immediately after having made them were significantly less likely to exhibit a judgmental error, such as the prominence effect. However, in the present experiment, this was not found.

In Experiment 4, an attempt was made to further increase the impact of the accountability manipulation by instructing subjects to justify and explain their evaluations as if they were physicians. Previous research has suggested that accountability may be regarded as inherent in the role of clinical professionals (Greenblatt, 1991; O'Neill, 1989). Nevertheless, this form of manipulation did not have any impact on the prominence effect.

It seems clear that the role of value ranges and accountability in judgment and choice is a topic that needs further investigation. Although we did not find any reliable effects of our accountability manipulations, the results suggest that the impact of accountability depends not only on the stimuli being used (Simonson & Nye, 1992) but also on how the manipulations themselves interplay with subjects' need to justify.

APPENDIX

TABLE A1
Stimuli for Experiment 1

Pair	Value ranges				
	05	10	15	20	25
1.	X_1 30 30 35	X_1 30 30 40	X_1 30 30 45	X_1 30 30 50	X_1 30 30 55
2.	35 X_2 30 35	40 X_2 30 40	45 X_2 30 45	50 X_2 30 50	55 X_2 30 55
3.	35 30 X_3 35	40 30 X_3 40	45 30 X_3 45	50 30 X_3 50	55 30 X_3 55
4.	35 30 30 X_4	40 30 30 X_4	45 30 30 X_4	50 30 30 X_4	55 30 30 X_4
5.	X_1 35 35 40	X_1 35 35 45	X_1 35 35 50	X_1 35 35 55	X_1 35 35 60
6.	40 X_2 35 40	45 X_2 35 45	50 X_2 35 50	55 X_2 35 55	60 X_2 35 60
7.	40 35 X_3 40	45 35 X_3 45	50 35 X_3 50	55 35 X_3 55	60 35 X_3 60
8.	40 35 35 X_4	45 35 35 X_4	50 35 35 X_4	55 35 35 X_4	60 35 35 X_4
9.	X_1 40 40 45	X_1 40 40 50	X_1 40 40 55	X_1 40 40 60	X_1 40 40 65
10.	45 X_2 40 45	50 X_2 40 50	55 X_2 40 55	60 X_2 40 60	65 X_2 40 65
11.	45 40 X_3 45	50 40 X_3 50	55 40 X_3 55	60 40 X_3 60	65 40 X_3 65
12.	45 40 40 X_4	50 40 40 X_4	55 40 40 X_4	60 40 40 X_4	65 40 40 X_4
13.	X_1 45 45 50	X_1 45 45 55	X_1 45 45 60	X_1 45 45 65	X_1 45 45 70
14.	50 X_2 45 50	55 X_2 45 55	60 X_2 45 60	65 X_2 45 65	70 X_2 45 70
15.	50 45 X_3 50	55 45 X_3 55	60 45 X_3 60	65 45 X_3 65	70 45 X_3 70
16.	50 45 45 X_4	55 45 45 X_4	60 45 45 X_4	65 45 45 X_4	70 45 45 X_4

Note. X_1 , Highest value on the prominent attribute missing; X_2 , lowest value on the prominent attribute missing; X_3 , lowest value on the nonprominent attribute missing; X_4 , highest value on the nonprominent attribute missing.

REFERENCES

- Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, **41**, 1–19.
- Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, **7**, 223–242.
- Chapman, G. B., & Johnson, E. J. (1995). Preference reversals in monetary and life expectancy evaluations. *Organizational Behavior and Human Decision Processes*, **62**, 300–317.
- Cialdini, R. B., Levy, A., Herman, C. P., & Evenbeck, S. (1973). Attitudinal politics: The strategy of moderation. *Journal of Personality and Social Psychology*, **39**, 752–766.
- Dahlstrand, U., & Montgomery, H. (1984). Information search and evaluative processes in decision making. *Acta Psychologica*, **56**, 113–123.
- 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.
- Greenblatt, M. (1991). Administrative psychiatry. *New Directions for Mental Health Services*, **49**, 5–17.
- Hagafors, R., & Brehmer, B. (1983). Does having to justify one's decisions change the nature of the judgement process? *Organizational Behavior and Human Performance*, **31**, 223–232.
- Hawkins, S. A. (1994). Information processing strategies in riskless preference reversals: The prominence effect. *Organizational Behavior and Human Decision Processes*, **59**, 1–26.
- Hershey, J., & Schoemaker, P. (1985). Probability versus certainty equivalence methods in utility assessments: Are they equivalent? *Management Science*, **31**, 1213–1231.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *The American Psychologist*, **39**, 341–350.
- Lindberg, E., Gärling, T., & Montgomery, H. (1989). Differential predictability of preferences and choice. *Journal of Behavioral Decision Making*, **2**, 205–219.
- Montgomery, H. (1983). Decision rules and the search for a dominance structure: Towards a process model of decision making. In P. C. Humphreys, O. Svenson, & A. Vari (Eds.), *Analyzing and aiding decision processes* (pp. 343–369). Amsterdam, North Holland Budapest: Academic Press Hungary.
- Montgomery, H., Selart, M., Gärling, T., & Lindberg, E. (1994). The judgement-choice discrepancy: Noncompatibility or restructuring? *Journal of Behavioral Decision Making*, **7**, 145–155.
- O'Neill, P. T. (1989). Responsible to whom? Responsible to what? Some ethical issues in community intervention. *American Journal of Community Psychology*, **17**, 323–341.
- Payne, J. W., Bettman, J. R., Coupey, E., & Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, **80**, 107–141.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, **43**, 87–131.
- Pruitt, D. (1981). *Negotiation behavior*. New York: Academic.
- Schkade, D. A., & Johnson, E. J. (1992). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, **44**, 203–231.
- Schkade, D. A., & Kleinmuntz, D. N. (1994). Information displays and choice processes: Differential effects of organization, form, and sequence. *Organizational Behavior and Human Decision Processes*, **57**, 319–337.
- Selart, M. (1994a). *Preference reversals in judgment and choice: The prominence effect*. Unpublished doctoral dissertation, Göteborg University, Sweden.
- Selart, M. (1994b). *Cognitive restructuring and violations of procedure invariance in preference measurement*. Paper presented at the 7th international conference on the foundations and applications of utility, risk, and decision theory (FURVII), Norwegian School of Management, Sandvika, Norway.
- Selart, M., Montgomery, H., Romanus, J., & Gärling, T. (1994). Violations of procedure invariance in preference measurement: Cognitive explanations. *European Journal of Cognitive Psychology*, **6**, 417–436.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, **16**, 158–174.
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, **51**, 416–446.
- 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.
- Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and postdecision processes. *Acta Psychologica*, **80**, 143–148.
- Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, **46**, 285–292.
- Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, **48**, 227–236.
- Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol 25.) New York: Academic.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327–352.
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, **95**, 371–384.
- Westenberg, M. R. M., & Koele, P. (1990). Response modes and decision strategies. In K. Borchering, O. L. Larichev, & D. Messick (Eds.), *Contemporary issues in decision making* (pp. 159–170). Amsterdam: North-Holland.
- Westenberg, M. R. M., & Koele, P. (1992). Response modes and decision processes. *Acta Psychologica*, **80**, 169–184.

Received: April 7, 1995