To appear in R. M. Hogarth (Ed.), Insights in decision making: Theory and applications. (A tribute to the late Hillel J. Einhorn.)

Compatibility Effects in Judgment and Choice

Paul Slovic, Decision Research and The University of Oregon

Dale Griffin, University of Waterloo

Amos Tversky, Stanford University

Support for this research was provided by the Air Force Office of Scientific Research and by NSF Grant No. SES 8712145 to Decision Research. This chapter has benefited from the comments of Robyn Dawes, Gregory Fischer, Robin Hogarth, Eric Johnson, and Daniel Kahneman.

Abstract

We investigate the hypothesis that the weight of a stimulus attribute is enhanced by its compatibility with the response mode. The first section demonstrates compatibility effects in predictions of market value (Study 1) and course grades (Study 2). In each case, the weight of a stimulus attribute is greater when it matches the response scale than when it does not. The second section applies the compatibility principle to the study of choice, and investigates the hypothesis that preference reversals are caused by the fact that payoffs are weighted more heavily in pricing than in choice, as implied by compatibility. This account is supported in experiments on risky choice (Studies 3 and 5), and on time preferences (Study 4). Theoretical and practical implications of the compatibility hypothesis are discussed in the last section.

One of the main ideas that has emerged from behavioral decision research in the last two decades is a constructive conception of judgment and choice. According to this view, preferences and beliefs are actually constructed--not merely revealed--in the elicitation process. This conception is entailed by findings that normatively equivalent methods of elicitation often give rise to systematically different responses (see, e.g., Slovic, Fischhoff & Lichtenstein, 1982; Tversky, Sattath, & Slovic, 1988). To account for these data within a constructive framework, we seek explanatory principles that relate the characteristics of the task to the attributes of the objects under study. One such notion is the compatibility hypothesis, which states that the weight of a stimulus attribute is enhanced by its compatibility with the response.

The rationale for this hypothesis is twofold. First, non-compatibility between the input and the output requires additional mental operations, which often increase effort and error and may reduce impact. Second, a response mode may prime or focus attention on the compatible features of the stimulus. Common features, for example, are weighted more heavily in judgments of similarity than in judgments of dissimilarity, whereas distinctive features are weighted more heavily in judgments of dissimilarity (Tversky, 1977). Consequently, entities with many common features and many distinctive features (e.g., East Germany and West Germany) are judged as both more similar to each other and as more different from each other than entities with relatively fewer common features and fewer distinctive features (e.g., Sri Lanka and Nepal).

The significance of the compatibility between input and output has long been recognized by students of human performance. Engineering psychologists have discovered that responses to visual displays of information, such as an instrument panel, will be faster and more accurate if the response structure is compatible with the arrangement of the stimuli (Fitts & Seeger, 1953; Wickens, 1984). For example, the response to a pair of lights will be faster and more accurate if the left light is assigned to the left key and the right light to the right key. Similarly, a square array of four burners on a stove is easier to control with a matching square array of knobs than with a linear array. The concept of compatibility has been extended beyond spatial organization. The reaction time to a stimulus light is faster with a pointing response than with a vocal response, but the vocal response is faster than pointing if the stimulus is presented in an auditory mode (Brainard, Irby, Fitts, and Alluisi 1962).

The present chapter investigates the role of compatibility in judgment and choice. As in the study of perceptual-motor performance, we do not have an independent procedure for assessing the compatibility between stimulus elements and response modes. This hinders the development of a general theory, but it does not render the concept meaningless or circular, provided compatibility can be experimentally manipulated. For example, it seems reasonable to assume that a turn signal in which a left movement indicates a left turn and a right movement indicates a right turn is more compatible than the opposite design. By comparing people's performance with the two turn signals, it is possible to test whether the more compatible design yields better performance. Similarly, it seems reasonable to assume that the monetary payoffs of a bet are more compatible with pricing than with choice, because both the payoffs and the prices are expressed in dollars. By comparing choice and pricing, therefore, we can test the hypothesis that the payoffs of a bet loom larger in pricing than in choice.

The research described in this chapter employs the notion of compatibility as a guiding principle that is translated into specific experimental hypotheses. In the first section we demonstrate compatibility effects in studies of prediction. The next section applies the compatibility hypothesis to the analysis of preference reversals in both risky and riskless choice. Theoretical and practical implications of the findings are addressed in the final section.

Prediction

Study 1: Prediction of Market Value

In all the following studies, subjects were either undergraduate students at Stanford University participating for course credit, or students at the University of Oregon who responded to an ad in the student newspaper and were paid for their participation. In our first study, seventy-seven Stanford students were presented with a list of twelve well-known U.S. companies taken from the 1987 *Business Week* Top 100. For each company, students were given two items of information: i) 1986 *market value* (i.e., the total value of the outstanding shares in billions of dollars), and ii) 1987 *profit standing* (i.e., the rank of the company in terms of its 1987 earnings among the Top 100); see Table 1. Half of the subjects were asked to predict 1987 market value (in billions of dollars). They were informed that the highest market value in 1987 was 68.2 billion dollars and the lowest (among the Top 100) was 5.1 billion dollars, so their predictions should fall within that range. The remaining subjects were asked to predict each company's rank (from 1 to 100) in market value for 1987. Thus, both groups of subjects received identical information and predicted the same criterion, using a different response scale. Although the two response scales differ in units (dollar versus rank) and direction (low rank means high market value), the two dependent variables should yield the same ordering of the twelve companies. To encourage careful consideration, a $75 prize was offered for the person whose predictions most nearly matched the actual values. The mean predicted values for each group are presented in Table 1 along with the actual values.

Insert Table 1 about here

The compatibility hypothesis states that a predictor will be weighted more heavily when it matches the response scale than when it does not. That is, 1986 market value in dollars should be weighted more heavily by the subjects who predict in dollars than by those who predict in rank. By the same token, 1987 profit rank should be weighted more heavily by the subjects who predict in rank than by those who predict in dollars. To investigate this hypothesis, we (i) correlated the criteria with the predictors, (ii) estimated the relative weights of the two predictors, and (iii) devised a statistical test based on reversals of order.

The product-moment correlations of d with D and R were .93 and .77, respectively, whereas the correlations of r with D and R were .74 and .94. Thus, the correlation between the matched variables was higher than that between the nonmatched variables. It is instructive to examine the compatibility effect in terms of the relative weights of the two predictors in a multiple regression equation. These values can be computed directly, or derived from the correlations between the predictors and the criterion together with the correlation between the predictors. (To make the regression weights positive, the ranking order was reversed). The multiple regres-

Table 1

Financial information for the twelve companies used to test the compatibility hypothesis with the respective mean predictions (actual outcome values in parenthesis).

		Predictors		*Criteria*	
		D	R	d	r
#	Company	1986 Market Value in billions	1987 Profit Rank (1 to 100)	1987 Market Value in billions	1987 Market Rank (1 to 100)
1	Chevron Corp.	$18.0	26	$21.3 (16.2)	30 (15)
2	H. J. Heinz	$6.2	75	$7.3 (5.6)	70 (84)
3	Coca-Cola	$18.1	31	$21.6 (14.8)	31 (17)
4	Westinghouse	$9.3	36	$12.9 (7.4)	44 (51)
5	Dow Chemical	$15.5	16	$20.5 (16.9)	26 (13)
6	Xerox	$7.1	54	$9.5 (5.7)	53 (82)
7	Chrysler	$8.2	12	$15.5 (5.5)	32 (90)
8	Kraft	$8.4	74	$9.0 (7.3)	64 (53)
9	Hewlett-Packard	$14.7	39	$17.4 (15.5)	42 (16)
10	Procter & Gamble	$15.6	63	$16.3 (13.9)	47 (25)
11	Kodak	$16.9	20	$20.9 (13.7)	27 (26)
12	Johnson & Johnson	$15.5	35	$18.2 (14.7)	36 (18)

sions for both dollars and ranks fit the average data very well with multiple correlations of .99. Let d_i and r_i denote the mean observed predictions of 1987 dollar value and rank, respectively, for a company whose 1986 dollar value is D_i and whose 1987 profit rank is R_i. The multiple regression equations, then, take the form

$$d_i = \alpha_d D_i + \beta_d R_i$$

$$r_i = \alpha_r D_i + \beta_r R_i,$$

when the independent variables are expressed in standardized units. Thus, α_d and α_r are the regression weights for the 1986 market value (D_i) estimated respectively from the predicted dollars and ranks. Similarly, β_d and β_r are the corresponding weights for the second predictor, 1987 profit rank. The relative weights for the first predictor in each of the two response modes are

$$A_d = \alpha_d/(\alpha_d+\beta_d)$$

and

$$A_r = \alpha_r/(\alpha_r+\beta_r).$$

These values measure the relative contribution of D_i in the prediction of dollars and rank, respectively. If the weighting of the dimensions is independent of the response scale, A_d and A_r are expected to be equal, except for minor perturbations due to a nonlinear relation between d and r. As we shall argue next, the compatibility hypothesis implies $A_d > A_r$. Note that A_d is the relative weight of the 1986 market value in dollars, estimated from the prediction of dollars, whereas A_r is the relative weight of the same variable estimated from the prediction of rank. The first index reflects the impact of D_i in a compatible condition (i.e., when the predictions are made in dollars), while A_r reflects the impact of D_i in the less compatible condition (i.e., when the predictions are made in ranks). If the compatibility between the predictor and the criterion enhances the weight of that variable then A_d should exceed A_r.

The values estimated from the regression equations were A_d=.64 and A_r=.32, in accord with the compatibility hypothesis. Thus, D_i was weighted more than R_i in the prediction of dollars, whereas R_i was weighted more than D_i in the prediction of rank. Moreover, each predictor was weighted about twice as much in the compatible condition than in the non-compatible condition. When interpreting the relative weights, here and in later studies, we should keep in mind (i) that they are based on aggregate data, (ii) that the predictors (D and R) are correlated, and (iii) that the relation between the two criteria (d and r) should be monotone but not necessarily linear. Although these factors do not account for the discrepancy between A_d and A_r, it is desirable to obtain a purely ordinal test of the compatibility hypothesis within the data of each subject that is not open to these objections. The following analysis of order reversals provides a basis for such a test.

The change in the relative weights induced by the response mode could produce reversals in the order of the predictions. In the present study, there were 21 pairs of companies (i,j) in which $D_i > D_j$ and $R_j > R_i$. If D is weighted more heavily than R in the subject's prediction of dollars, and R is weighted more heavily than D in the subject's prediction of rank, we would expect $d_i > d_j$ and $r_j > r_i$. The data confirmed this hypothesis. Subjects who predicted dollars favored the company with the higher D 72% of the time, whereas subjects who predicted rank favored the company with the higher D only 39% of the time. (Ties were excluded from this analysis.) This difference is highly significant ($p < .001$). Note that the subjects did not directly compare the companies; the ordering was inferred from their predictions.

Insert Figure 1 about here

Figure 1 provides a graphical summary of the stimuli and the data. Each of the twelve companies is represented as a point in the D × R plane. Each regression equation defines a set of parallel equal-value lines. The points on any given line are the values of the two predictors that give rise to the same predicted value of the criterion. For the prediction of dollar, for instance, each equal-value line is the set of points for which $\alpha_d D_i + \beta_d R_i$ is a constant. The two prediction lines (d and r) are perpendicular to the equal-value lines for the two criteria. Hence, the predicted order of the companies is given by the order of their projections, denoted by notches. The slopes of the prediction lines are the weight ratios, α_d/β_d and α_r/β_r, of D to R, estimated from d and r, respectively. It is evident from the figure that, in accord with the compatibility hypothesis, the two criteria induced different orders of the twelve companies. For example, the predicted market value of Chevron (#1) is higher than that of Dow Chemical (#5), but the latter is assigned a higher rank than the former.

Study 2: Prediction of academic performance

Our second test of the compatibility hypothesis involves the prediction of a student's grade in a course. Two hundred and fifty-eight subjects from the University of Oregon predicted the performance of 10 target students in a History course on the basis of the students' performance in two other courses: English Literature and Philosophy. For each of the 10 targets, the subjects were given a letter grade (from A+ to D) in one course, and a class rank (from 1 to 100)

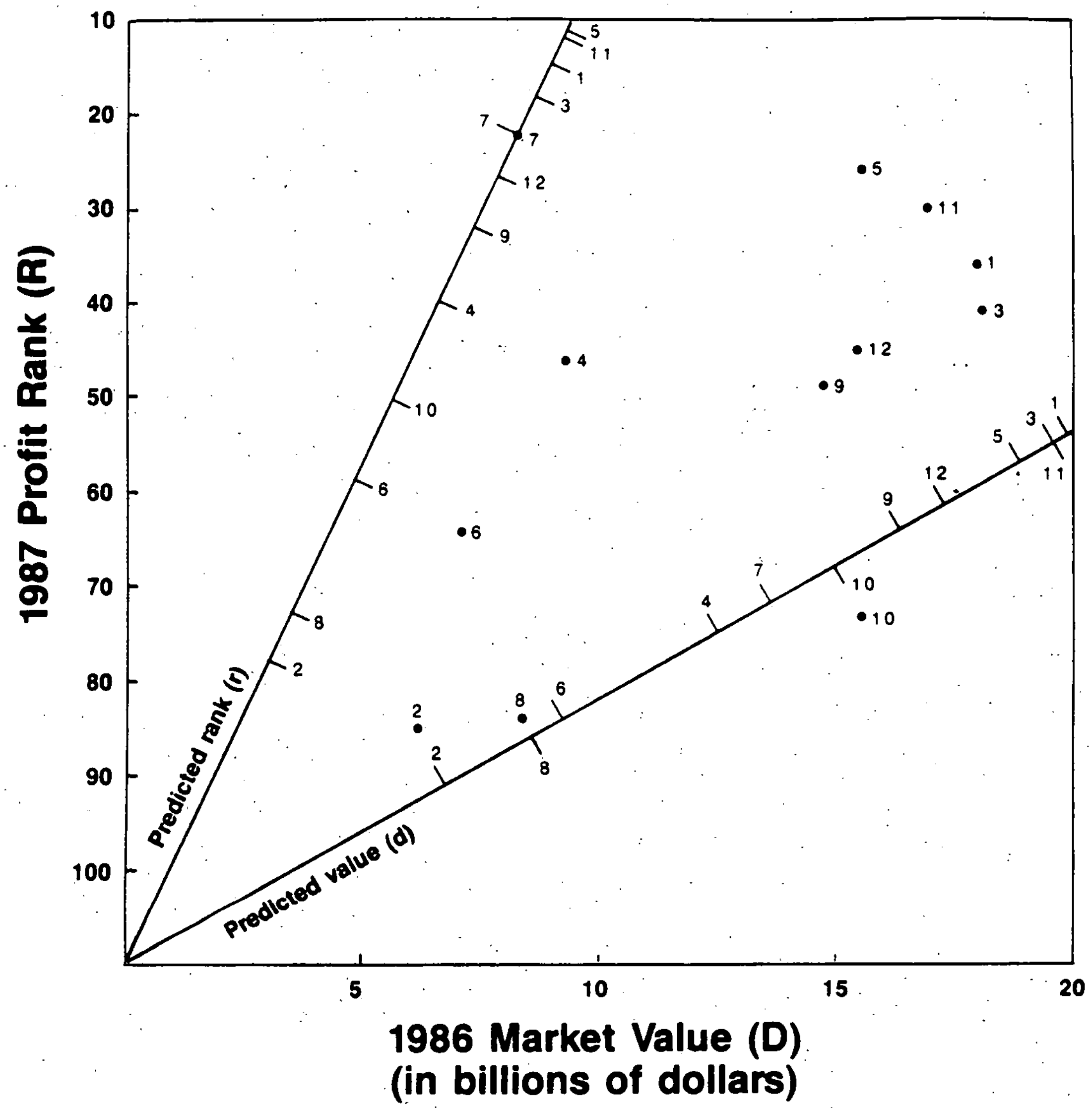

Figure 1. A graphical summary of Study 1. The dots represent the twelve companies. The slopes of the d and r lines correspond to the weight ratios, α_d/β_d and α_r/β_r, of D to R, in the two prediction tasks.

in the other course. One-half of the subjects predicted the students' grade in History whereas the other half predicted the students' class rank in History. Each of the four combinations of performance measures (grade/rank) and courses (Literature/Philosophy) was presented to a different group of subjects. The description of the 10 hypothetical students is presented in Table 2 along with the mean predictions of grade and rank, rounded to the nearest unit.

Insert Table 2 about here

The compatibility hypothesis implies that a given predictor (e.g., grade in Philosophy) will be given more weight when the criterion is expressed on the same scale (e.g., grade in History) than when it is expressed on a different scale (e.g., rank in History). The relative weight of grades to ranks, then, will be higher in the group that predicts grades than in the group that predicts ranks.

As in the previous study, we first correlated the criteria with the predictors. The (zero-order) correlations of g with G and R were .83 and .82, respectively, whereas the correlations of r with G and R were .70 and .91, in accord with the compatibility hypothesis. We next regressed the mean predictions of grades and ranks (displayed in Table 2) onto the two predictors. The letter grades were coded D=1, C-=2 ... A+=10. (To make the the regression weights positive, the ranking order was reversed). The multiple regressions for both grades and ranks fit the average data very well with multiple correlations of .99. Let g_i and r_i denote the mean observed predictions of grade and rank, respectively, for a student with a grade G_i in one course and a rank R_i in

Table 2

Academic performance of the ten hypothetical students used to test the compatibility hypothesis with the respective mean predictions.

	Predictors		*Criteria*	
	G	R	g	r
Student	Grade in Class 1 (A+ to D)	Rank in Class 2 (1 to 100)	Predicted Grade	Predicted Rank
1	B+	66th	C+	48
2	D	93rd	D	87
3	A	45th	B	33
4	C+	34th	B-	40
5	A+	6th	A	11
6	C-	54th	C	54
7	B	59th	B-	49
8	A-	72nd	B-	48
9	C	28th	B-	35
10	B-	41st	B-	38

the other course. There was no significant interaction between the scale (rank/grade) and the course (Literature/Philosophy), therefore, the data for the two courses were pooled. The multiple regression equations, then, take the form

$$g_i = \alpha_g G_i + \beta_g R_i$$

$$r_i = \alpha_r G_i + \beta_r R_i,$$

when the independent variables are expressed in standardized units. Thus, α_g and α_r are the regression weights for the grades (G_i) estimated respectively from the predicted grades and ranks. Similarly, β_g and β_r are the corresponding weights for the second predictor, class rank. The relative weights for the first predictor in each of the two response modes are

$$A_g = \alpha_g/(\alpha_g+\beta_g)$$

and

$$A_r = \alpha_r/(\alpha_r+\beta_r).$$

These values measure the relative contribution of G_i in the prediction of grade and rank, respectively. Because the grades and ranks are monotonically related, A_g and A_r should be approximately equal if the weighting of the dimensions is independent of the response scale. However, if the match between the predictor and the criterion enhances the weight of the more compatible predictor, then A_g should exceed A_r.

The values estimated from the regression equations were $A_g = .51$ and $A_r = .40$, in accord with the compatibility hypothesis. Thus, grade in Philosophy was weighted more heavily in the prediction of grade in History than in the prediction of rank in History. Similarly, rank in Philosophy was weighted more heavily in the prediction of rank in History than in the prediction of grade in History.

To obtain an ordinal test of the compatibility hypothesis within the data of each subject, we analyzed the reversals of order induced by the change in weights. There were 21 pairs of students (i,j) in which $G_i > G_j$ and $R_j > R_i$. If G is weighted more heavily than R in the prediction of grades, and R is weighted more heavily than G in the prediction of rank, we would expect $g_i > g_j$ and $r_j > r_i$. Indeed, subjects who predicted grades favored the student with the higher G 58% of the time, whereas subjects who predicted rank favored the student with the higher G only 42% of the time. (Ties were excluded from this analysis.) This difference is statistically significant ($p < .001$). Recall that subjects did not compare students directly; the ordering was inferred from their predictions. Figure 2 provides a graphical representation of these data.

Insert Figure 2 about here

The compatibility effects observed in the previous two studies may be mediated by a process of anchoring and adjustment. Subjects may use the score on the compatible variable (the attribute which matches the criterion) as an anchor, and then adjust this number upward or downward according to the value of the non-compatible variable. Because adjustments of an anchor are generally insufficient (Slovic & Lichtenstein, 1971; Tversky & Kahneman, 1974) the compatible attribute would be overweighted. An anchoring and adjustment process, therefore, provides a natural mechanism for generating compatibility effects. To test whether compatibility effects occur in the absence of anchoring, we replaced the prediction task described above with a choice task in which the subject is no longer required to make a numerical prediction that

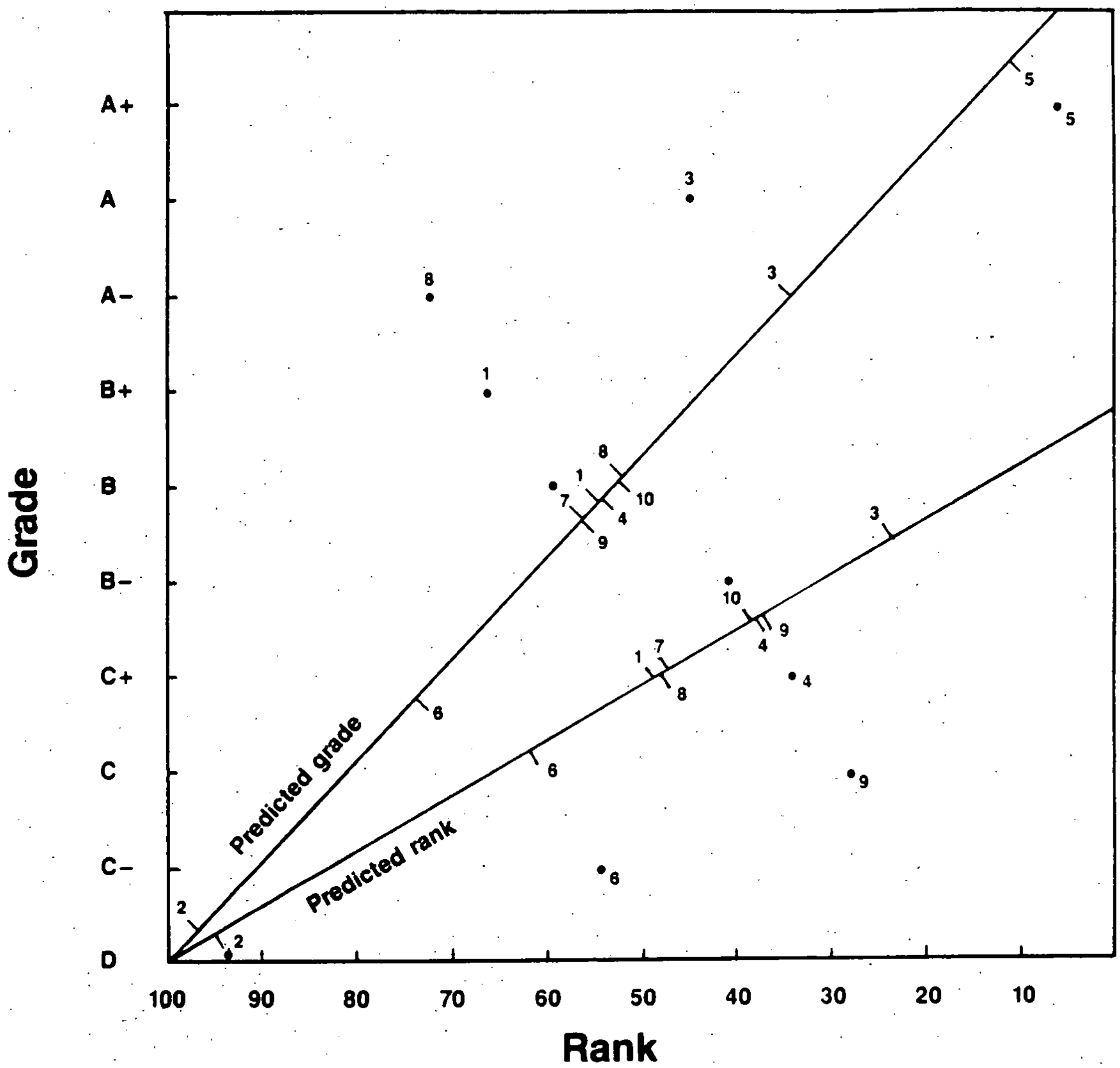

Figure 2. A graphical summary of Study 2. The dots represent the ten students. The slopes of the grade and rank lines correspond to the weight ratios, α_g/β_g and α_r/β_r, of grades to ranks in the two prediction tasks.

would invoke anchoring. The following study, then, investigates the compatibility hypothesis in a context in which anchoring and adjustment are unlikely to play a significant role.

Seventy-eight Stanford undergraduates were presented with 20 pairs of students taken from the list given in Table 2. In each pair, one student had a higher grade while the other had a higher rank. Half of the subjects were asked to predict, for each pair, which student would achieve a higher grade in History whereas the remaining subjects were asked to predict, for each pair, which student would achieve a higher rank in History. Because both groups were only asked to predict which of two students would do better in History, without making a numerical prediction, their tasks were virtually identical.

Nevertheless, the student with the higher grade was selected 56% of the time by the grade group and only 49% of the time by the rank group ($p < .05$), indicating that the compatibility effect is present even in a choice task that does not require a quantitative response and is, therefore, unlikely to involve an adjustment of a numerical anchor. The strategy of anchoring and adjustment, however, probably contributes to the compatibility effect observed in numerical predictions.

Preference

The previous section investigated compatibility effects in prediction and judgment. The present section is concerned with the role of compatibility in decision making in general, and preference reversals in particular. A reversal of preference is a pattern of choices in which normatively equivalent elicitation procedures give rise to inconsistent preferences. A well-known example of preference reversal was discovered by Lichtenstein and Slovic (1971; see also Slovic

& Lichtenstein, 1968). This phenomenon involves pairs of bets with comparable expected values: an H bet that offers a high probability of winning a relatively small amount of money (e.g., 35/36 chances to win $4) and an L bet that offers a low probability of winning a moderate amount of money (e.g., 11/36 chances to win $16). When offered a choice between such bets, most people choose the H bet over the L bet, but when asked to state the lowest selling price of each bet, the majority state a higher price for the L bet than for the H bet. In general, about half the subjects state prices that are inconsistent with their choices, thereby exhibiting a preference reversal, or PR for short. This pattern of preferences, which violates the standard theory of rational choice, has been observed in numerous experiments, including a study conducted on the floor of a Las Vegas casino (Lichtenstein & Slovic, 1973), and it persists even in the presence of monetary incentives designed to promote consistent responses (see, e.g., Grether & Plott, 1979; Slovic & Lichtenstein, 1983).

Let C_H and C_L denote, respectively, the cash equivalent (or minimum selling price) of the H bet and L bet, and let $>^*$ and $\approx$ denote strict preference and indifference, respectively. In this notation, PR is expressed as $H >^* L$, and $C_L > C_H$. Note that $>^*$ refers to preference between options, whereas $>$ refers to the ordering of cash amounts. (Naturally, $X > Y$ implies $X >^* Y$, that is, more money is preferred to less.)

It can be shown that PR violates either transitivity or procedure invariance, and possibly both (Tversky, Slovic & Kahneman, 1989). Procedure invariance states that choice and pricing yield the same ordering of options, that is, a bet B is preferred to a cash amount X if and only if the cash equivalent of B, C_B, exceeds X. In particular, $C_B = X$ whenever the decision maker is indifferent between playing the bet B and receiving the cash amount X. If procedure invariance holds, PR reduces to an intransitivity of the form

$$C_H \approx H >^* L \approx C_L >^* C_H.$$

On the other hand, >* may be transitive, in which case PR violates procedure invariance. This violation can be produced by either

(i) overpricing of L (i.e., $C_L >^* L$), or

(ii) underpricing of H (i.e., $H >^* C_H$).

It follows from this analysis that PR may be caused either by the intransitivity of >* or by a failure of procedure invariance that gives rise to a choice-pricing discrepancy. To investigate these possibilities, Tversky et al. (1989) extended the traditional design by including, in addition to the bets H and L, a cash amount X that is compared to both. By focusing on all cases in which $C_L > X > C_H$, it is possible to diagnose all PR patterns according to whether they imply an intransitive choice, an overpricing of L, an underpricing of H, or both overpricing of L and underpricing of H. Tversky et al (1989) applied this analysis to an extensive study of preference reversals, using 18 triples (H, L, X) that covered a wide range of probabilities and payoffs. The diagnostic analysis of the observed response patterns showed that the most important determinant of PR was the overpricing of L. Intransitive choice and the underpricing of H played a relatively minor role, each accounting for less than 10% of the total number of reversals.

The compatibility hypothesis offers a simple explanation for the overpricing of L bets. Because the selling price of a bet is expressed in dollars, we expect that the payoffs, which are expressed in the same units, will be weighted more heavily in pricing than in choice. To test this hypothesis, Tversky et al (1989) have employed a contingent weighting model in which the relative weight of an attribute varies with the method of elicitation. This analysis differs from the regression analysis discussed in the previous section in two important respects. First, the two attributes of a simple gamble, probability and payoff, combine multiplicatively rather than additively. Consequently, the multiple regression analysis was applied to the logarithms of the probabilities and the payoffs. Second, the analysis uses only the ordering of the bets by price and by choice. Specifically, assume that a bet $B = (P,X)$ is chosen over $B' = (P',X')$ iff

$$\log P + \alpha \log X > \log P' + \alpha \log X'.$$

Similarly, assume that B is priced higher than B' iff

$$\log P + \beta \log X > \log P' + \beta \log X'.$$

These relations are equivalent to the assumption that the ordering of bets according to both choice and pricing follows a multiplicative probability-value model with a power function for gains, and exponents α and β for choice and pricing, respectively. If the payoff of a bet looms larger in pricing than in choice, as implied by compatibility, β should exceed α.

To test this prediction, Tversky et al (1989) applied the above model to the data and estimated α and β separately for each subject. Note that a choice between an H-bet (P_H,X_H) and an L-bet (P_L,X_L) implies an inequality involving α. According to the above model, H is chosen over L iff

$$\log P_H + \alpha \log X_H > \log P_L + \alpha \log X_L,$$

or equivalently whenever

$$R = \log(P_H/P_L)/\log(X_L/X_H) > \alpha.$$

Any comparison of H_i and L_i, $i = 1, ..., 18$, gives rise to an inequality of the form $R_i > \alpha$ or $R_i < \alpha$. For each subject, a value of α was selected so as to minimize the average squared deviations between the model and the data. Specifically, for any subject and any pair of bets (H_i, L_i) define $x_i = 1$ if $H_i >^* L_i$ and $x_i = 0$ if $L_i >^* H_i$. A value of α was selected for each subject by minimizing the quadratic loss function

$$F(\alpha) = \sum_{i=1}^{18} f(\alpha, x_i) \text{ where}$$

$$f(\alpha, x_i) = \begin{cases} x_i(\alpha - R_i)^2 & \text{if } R_i < \alpha \\ (1 - x_i)(\alpha - R_i)^2 & \text{if } R_i > \alpha. \end{cases}$$

Exactly the same procedure was used to estimate β, except that the H_i, L_i pairs were ordered by their cash equivalents, excluding ties. In accord with the compatibility hypothesis, β exceeded α for 87% of the subjects (N=179) and the difference between them was significantly positive ($p<.001$). To evaluate the adequacy of the model, the logarithm of the prices were regressed against log P and log X, separately for each subject. The median value of the multiple correlation was .95, indicating that the model provided a reasonable fit for individual data.

It should be noted that the contingent-weighting model (with $\beta > \alpha$) implies overpricing of both H and L bets. It can be shown that the predicted effect, however, is substantial for L bets and negligible for H bets. More specifically, let Y_c and Y_p, respectively, be the cash amounts that are equivalent to the bet (P,X) in choice and in pricing. It follows from the model that the discrepancy between choice and pricing, measured by $\log(Y_p/Y_c)$, is proportional to log P. It vanishes when P approaches 1, and it is large when P is small. For example, the overpricing effect implied by the model is 20 times larger when the probability of winning (P) is .1 than when it is .9. In general, P is above .9 for H bets and below .5 for L bets. The contingent-weighting model, therefore, explains the major cause of preference reversal, namely, the overpricing of L bets. Additional hypotheses are required to explain second-order effects, such as the occasional intransitivities and the slight underpricing of H bets. In the remainder of this section, we test other implications of the compatibility hypothesis in both risky and riskless choice.

Study 3: Monetary vs. nonmonetary outcomes

If preference reversals are due primarily to the compatibility of prices and payoffs, their frequency should be substantially reduced when the outcomes of the bets are not expressed in monetary terms. To test this prediction, we constructed six pairs of H and L bets, three with monetary outcomes (as in the usual PR studies) and three with nonmonetary outcomes. Two hundred and forty-eight students from the University of Oregon participated in this study. Half of the subjects first chose between all six pairs of bets and later assigned a cash equivalent to each bet. The other half of the subjects performed these tasks in the opposite order. There was no significant order effect, therefore, the data for the two groups were combined. Table 3 presents the entire set of twelve bets and the percentage of subjects who preferred the H bet over the L bet (H > L), the percentage of subjects who assigned a higher cash equivalent to H than to L ($C_H > C_L$), and the percentage of preference reversals (PR).

Insert Table 3 about here

The data show that the percentage of choices of H over L was roughly the same in the monetary and the nonmonetary bets (63% vs. 66%), but the percentage of cases in which C_H exceeds C_L was substantially smaller in the monetary than in the nonmonetary bets (33% vs. 54%). Consequently, the overall incidence of predicted preference reversal decreased significantly from 41% to 24% ($p<.01$). Naturally, the pricing response is more compatible with monetary payoffs than with nonmonetary payoffs. Hence, the observed reduction in preference reversal with nonmonetary outcomes underscores the role of compatibility in the evaluation of options. Because even the nonmonetary payoffs can be evaluated in monetary terms, albeit with some difficulty, we do not expect the complete elimination of preference reversals in this case.

Study 4: Time Preferences

The compatibility hypothesis entails that preference reversals should not be restricted to risky choice and they should also be found in riskless options. The present study investigates this hypothesis using delayed payoffs that differ in size and length of delay (see Tversky et al, 1989). Consider a delayed payoff of the form (X, T) that offers a payment of X dollars, T years from now. Table 4 presents four pairs of options that consist of a long-term prospect L (e.g.,

Table 3

The monetary and nonmonetary bets used to test the compatibility hypothesis with the respective percentage of preferences.

		$H >^* L$	$C_H > C_L$	PR
	Monetary Bets			
1.	H: .94 to win $3 L: .50 to win $6.50	57	26	42
2.	H: .86 to win $7.50 L: .39 to win $17	69	21	51
3.	H: .81 to win $16 L: .19 to win $56	63	51	29
	Mean	63	33	41
	Nonmonetary Bets			
4.	H: .89 to win a one-week pass good at all movie theatres in town. L: .33 to win a one-month pass good at all movie theatres in town.	65	46	30
5.	H: .92 to win an all-expenses-paid weekend at an Oregon coastal resort. L: .08 to win a one-week all-expenses-paid trip to Hawaii.	72	56	25
6.	H: .92 to win a one-week pass good at all movie theatres in town. L: .31 to win dinner for two at a very good restaurant.	62	60	16
	Mean	66	54	24

$2500, 5 years from now), and a short-term prospect S (e.g., $1600, 1 1/2 years from now).

One hundred and sixty-nine students from the University of Oregon participated in a study of choice between delayed payoffs. One-half of the subjects first chose between S and L in each pair and later priced all eight options by stating "the smallest immediate cash payment for which they would be willing to exchange the delayed payment". The other subjects performed the choice and pricing tasks in the opposite order. There were no systematic differences between the groups, so their data were combined.

Insert Table 4 about here

Table 4 presents the four pairs of options employed in this study. The table also includes, for each pair, the percentage of subjects who chose S over L (S >* L), the percentage of subjects who priced S above L ($C_S > C_L$), and the percentage of PR patterns (S >* L and $C_L > C_S$). Because both the given payoffs and the stated prices are expressed in dollars, the compatibility hypothesis implies that the payoffs will be weighted more heavily in pricing than in choice. As a consequence, the preference for the short-term option (S) over the long-term option (L) should be greater in choice than in pricing. Table 4 confirms this prediction. Overall, S was chosen over L 74% of the time, but S was priced higher than L only 25% of the time, yielding 52% preference reversals, as compared with 3% reversals in the opposite direction. The application of the diagnostic analysis described earlier revealed that, as in the case of choice between simple bets, the major determinant of preference reversal was overpricing of the long-term option, as

Table 4

The options used in Study 5 and the respective percentage of preferences. The pair (X,T) denotes the option of receiving $X, T years from now.

S	L	S >* L	$C_S > C_L$	PR
(1600, 1 1/2)	(2500, 5)	57	12	49
(1600, 1/ 1/2)	(3550, 10)	72	19	56
(2500, 5)	(3550, 10)	83	29	57
(1525, 1/2)	(1900, 2 1/2)	83	40	46
Mean		74	25	52

suggested by compatibility (Tversky et al, 1989).

In the pricing task each option is evaluated singly whereas choice involves a direct comparison between options. The standard demonstrations of PR, therefore, are consistent with the alternative hypothesis that payoffs are weighted more heavily in a singular than in a comparative evaluation. To test this hypothesis against compatibility, we replicated the above study on a new group of 184 students from the University of Oregon, with one change. Instead of pricing the options, the subjects were asked to rate the attractiveness of each option on a scale from 0 (not at all attractive) to 20 (extremely attractive). If PR is controlled, in part at least, by the nature of the task (singular vs. comparative) we should expect L to be more popular in rating than in choice. On the other hand, if PR is produced by scale compatibility, there is no obvious reason why rating should differ from choice. Indeed, no discrepancy between choice and rating was observed. Overall, S was chosen over L 75% of the time (as in the original study) and the rating of S exceeded the rating of L in 76% of the cases. Only 11% of the patterns exhibited PR between choice and rating as compared to 52% between choice and pricing.

Study 3 showed that the use of nonmonetary prizes greatly reduced the amount of preference reversal whereas Study 4 demonstrated substantial preference reversal in the absence of risk. Evidently, preference reversals are controlled primarily by the compatibility between the price and the payoffs, regardless of the presence or absence of risk.

Study 5: Matching vs. Pricing

In addition to pricing and choice, options can be evaluated through a matching procedure in which a decision maker is required to fill in a missing value so as to equate a pair of options. Considerations of compatibility suggest that the attribute on which the match is made will be overweighted relative to another attribute. This hypothesis is tested in the following study, using 12 pairs of H and L bets, displayed in Table 5. In each pair, one value -- either a probability or a payoff -- was missing, and the subjects were asked to set the missing value so they would be indifferent between the two bets. Consider, for example, the bets H = (33/36; $50) and L = (18/36; $125). If we replace the 18/36 probability in L by a question mark, the subject is asked in effect "what chance to win $125 is equally attractive as a 33/36 chance to win $50?" The value set by the subject implies a preference between the original bets. If the value exceeds 1/2, we infer that the subject prefers H to L, and if the value is less than 1/2 we reach the opposite conclusion. Using all four components as missing values, we can infer the preferences from matching either the probability or the payoff of each bet. If the compatibility hypothesis applies to matching, then the attribute on which the match is made will be overweighted relative to the other attribute. As a consequence, the inferred percentage of preferences for H over L should be higher for probability matches than for payoff matches.

Two hundred subjects from the University of Oregon participated in this study. Each subject saw 12 pairs, each consisting of a high probability bet (H) and a low probability bet (L). Six of these pairs consisted of bets with relatively small payoffs; the other six pairs consisted of bets with large payoffs, constructed by multiplying the payoffs in the first six pairs by a factor of 25 (see Table 5). Each pair of bets was evaluated in four ways: direct choice, pricing of each bet individually, matching by providing a missing payoff, and matching by providing a missing probability. Every subject performed both choice and pricing tasks, and matched either probabilities or payoffs (no subject matched both probabilities and payoffs). The order in which these tasks were performed was counterbalanced.

Insert Table 5 about here

The dependent variable of interest is the percentage of responses favoring the H bet over the L bet. These values are presented in Table 5 for all four tasks. Note that these percentages are directly observed in the choice task and inferred from the stated prices and the probability and payoff matches in the other tasks. Under procedure invariance, all these values should coincide. The overall means showed that the tendency to favor the H bet over the L bet was highest in choice (76%) and in probability matching (73%), and substantially smaller in payoff matching (47%) and in pricing (37%). These results demonstrate two types of preference reversals: i) choice versus pricing, and ii) probability matching versus payoff matching.

i) Choice versus pricing. The comparison of the results of choice and pricing in Table 5 reveals the familiar PR pattern. Subjects preferred the H bet but assigned a higher cash equivalent to the L bet. As was demonstrated earlier, this effect is due primarily to the overpricing of L bets implied by compatibility.

Table 5

Percentage of responses favoring the H bet over the L bet for four different elicitation procedures.

H		L	Choice	Probability Matching	Payoff Matching	Pricing
Small Bets:						
(35/36,$4)	or	(11/36,$16)	80	79	54	29
(29/36,$2)	or	(7/36,$9)	75	62	44	26
(34/36,$3)	or	(18/36,$6.5)	73	76	70	39
(32/36,$4)	or	(4/36,$40)	69	70	26	42
(34/36,$2.5)	or	(14/36,$8.5)	71	80	43	22
(33/36,$2)	or	(18/36,$5)	56	66	69	18
Mean			71	72	50	29
Large Bets:						
(35/36,$100)	or	(11/36,$400)	88	76	69	65
(29/36,$50)	or	(7/36,$225)	83	64	31	55
(34/36,$75)	or	(18/36,$160)	77	79	65	55
(32/36,$100)	or	(4/36,$1,000)	84	68	28	61
(34/36,$65)	or	(14/36,$210)	78	80	36	57
(33/36,$50)	or	(18/36,$125)	68	75	58	46
Mean			80	74	48	56
Overall mean			76	73	49	37

ii) Probability matching versus payoff matching. The major new result of this study concerns the discrepancy between probability matching and payoff matching. By compatibility, the dimension on which the match is made should be overweighted relative to the other dimension. Probability matching, therefore, should favor the H bet, whereas payoff matching should favor the L bet. Indeed, the tendency to favor the H bet over the L bet was much more pronounced in probability matching than in payoff matching.

Table 5 contains two other comparisons of interest: pricing versus payoff matching, and choice versus matching. Although the pricing of a bet can be viewed as a special case of payoff matching in which the matched bet has P = 1, it appears that the monetary dimension looms even larger in pricing than in payoff matching. This conclusion, however, may not be generally valid, since it holds for the small but not the large bets.

Finally, the least expected feature of Table 5 concerns the relation between choice and matching. If, relative to choice, probability matching biases the responses in favor of the H bets whereas payoff matching biases the responses in favor of the L bets, then the choice data should lie between the two matching conditions. The finding that the tendency to favor the H bet is about the same in direct choice and in probability matching suggests that an additional effect beyond scale compatibility is involved.

The missing factor, we propose, is the prominence effect demonstrated by Tversky et al (1988). In an extensive study of preference, these investigators showed that the more important attribute of an option is weighted more heavily in choice than in matching. In other words, the choice ordering is more lexicographic than that induced by matching. We have originally interpreted PR in terms of compatibility rather than prominence (Tversky et al., 1988), because we saw no a priori reason to hypothesize that probability is more important than money. The results of Study 5, however, forced us to reconsider the hypothesis that probability is more prominent than money, which is further supported by the finding that the rating of bets is dominated by probability (see Goldstein & Einhorn, 1987; Slovic & Lichtenstein, 1968; Tversky et. al, 1988). It appears to us now that the data of Table 5 represent the combination of two effects: a compatibility effect that is responsible for the difference between probability matching and payoff matching (including pricing), and a prominence effect that contributes to the relative attractiveness of H bets in choice. This account is illustrated in Table 6 which characterizes each of the four elicitation procedures in terms of their compatibility and prominence effects.

Insert Table 6 about here

Let us examine first the columns of Table 6, which represent the effects of the compatibility factor. Recall that the probability matching procedure enhances the significance of P and thereby favors the H bet. Analogously, the compatibility of the payoff matching and pricing procedures with the monetary outcomes enhances the significance of the payoffs and thereby favors the L bet. The choice procedure, however, is neutral with respect to the compatibility factor, hence it is expected to lie between the two matching procedures-- if compatibility alone were involved. Now consider the rows of Table 6. In terms of the prominence factor, the more important dimension (i.e., probability) is expected to loom larger in choice than in either matching procedure. Thus, the tendency to choose the H bet should be greater in choice than in

Table 6

Compatibility and prominence effects for four elicitation procedures.

Prominence effect favors		Compatibility effect favors: H	Neither	L
	H		Choice	
	Neither	Probability Matching		Payoff Matching, Pricing

matching, if prominence alone were involved. Table 5 suggests that both compatibility and prominence are present in the data. The finding that choice and probability matching yield similar results suggests that the two effects have roughly the same impact. It follows from this analysis that compatibility and prominence contribute jointly to the discrepancy between choice and pricing, which may help explain both the size and the robustness of the standard preference reversal. It is noteworthy that each of these effects has been established independently. The demonstrations of compatibility reported in the first part of this paper do not involve prominence, and the prominence effects demonstrated by Tversky et al (1988) do not depend on scale compatibility.

Discussion

Although the notion of compatibility has long been suggested as a possible cause of elicitation effects (see, e.g., Lichtenstein & Slovic, 1971; Slovic & MacPhillamy, 1974), this hypothesis has not heretofore been tested directly. The present investigations tested several implications of the compatibility hypothesis in studies of prediction and preference. In each of these studies, enhancing the compatibility between a stimulus attribute and the response mode led to increased weighting of that attribute. These findings indicate that compatibility plays an important role in judgment and choice. At the same time it is evident that this concept requires further theoretical analysis and empirical investigation. Implications of the present work and directions for future studies are discussed below.

The testing and application of the compatibility principle require auxiliary hypotheses about the characteristics of a stimulus attribute that make it more or less compatible with a given response mode. Many features of stimulus attributes and response scales could enhance their compatibility. These include the use of the same units (e.g., grades, ranks), the direction of relationships (e.g., whether the correlations between input and output variables are positive or negative), and the numerical correspondence (e.g., similarity) between the values of input and output variables. Although we do not have a general procedure for assessing compatibility, there are many situations in which the compatibility ordering could be assumed with a fair degree of confidence. For example, it seems evident that the prediction of market value in dollars is more compatible with a predictor expressed in dollars than with a predictor expressed in ranks. The same situation exists in the domain of perceptual-motor performance. There is no general theory for assessing the compatibility between an information display and a control panel, yet it is evident that some input-output configurations are much more compatible than others and therefore yield better performance.

Further evidence for compatibility effects in risky choice has been reported by Schkade and Johnson (1988). Using a computer-controlled experiment in which the subject can see only one component of each bet at a time, the investigators were able to measure the amount of time spent by each subject looking at probabilities and at payoffs. Their results showed that the percentage of time spent on payoffs was significantly greater in pricing than in choice. Furthermore, this pattern was particularly pronounced when the subjects produced preference reversals, and it vanished when the subjects produced consistent responses. The conclusion that subjects attend to the payoffs in pricing more than in choice supports the hypothesis that subjects focus their attention on the stimulus components that are most compatible with the response mode. This finding is also consistent with the hypothesis that, in choice between bets, probability is perceived as more important than payoff.

In a second experiment, Schkade and Johnson (1988) compared the pricing of bets to their rating on a 100 point scale. The participants in this study expressed the ratings and the prices using an adjustable pointer. The authors observed that both the initial and the final settings of the pointer were higher for the L bet than for the H bet in pricing, and higher for the H bet than for the L bet in rating. The authors attribute the reversal of preference observed in this task to an insufficient adjustment (Slovic & Lichtenstein, 1971; Tversky & Kahneman, 1974) of the self-generated anchors. The productions of these anchors, however, appears to be governed by compatibility. Note that the response scale in the pricing task ranges from 0 to the positive payoff, whereas the range of the rating scale (0 to 100) matches the probability scale. By compatibility, the payoff is expected to loom larger in pricing than in rating, and the probability is expected to loom larger in rating than in pricing. The notion that the bounded rating scale is more compatible with probability than with money, supported by the process data of Schkade and Johnson, may explain the finding (Goldstein & Einhorn, 1987) that the preference for the H bet over the L bet is stronger in rating than in choice, despite the procedural similarity between rating and pricing. An alternative explanation of this result that attributes PR to the mapping of subjective value onto the response scale rather than to the compatibility between stimulus components and response modes was proposed by Goldstein and Einhorn (1987). Their model can accommodate reversals of preferences, but it does not predict the variety of compatibility effects described in the present paper.

Recent results reported by Delquie´ and de Neufville (1988) are also consistent with the compatibility hypothesis. These authors employed a double-matching procedure, devised by Hershey and Schoemaker (1985), in which subjects first determine the missing value (e.g., the probability of winning) of an option that would make it equivalent to a second option. Later, the subjects are presented with the option they constructed and they now have to determine the missing value (e.g., the payoff) of the second option that would make the two options equally attractive. If procedure invariance holds, the latter match should coincide with the given value of the second option. Using both risky and riskless options, Delquie´ and de Neufville found systematic violations of procedure invariance, which imply that the matched attribute is weighted more heavily than the other attribute -- as predicted by compatibility. These findings confirm, in a double-matching design, the conclusion of Experiment 5 and of Tversky et. al (1988) that were based on a choice-matching design. Another difference between the designs is that Delquie´ and de Neufville used a series of converging choices, rather than direct matching, to establish indifference. Because the subjects in these experiments are not required to make numerical responses, the results cannot be readily attributable to an insufficient adjustment of an anchor. As we have noted earlier, this heuristic could generate compatibility effects, but these occur even in tasks that do not require a numerical response and are, therefore, unlikely to evoke a process of anchoring and adjustment.

The compatibility notion discussed in this paper concerns the correspondence between the scales in which the inputs and outputs are expressed. In a previous paper (Tversky et al., 1988), we have explored a more abstract notion of compatibility that was later called "strategy compatibility" by Fischer and Hawkins (1988). To introduce this concept, we distinguished between qualitative and quantitative choice strategies. Qualitative strategies (e.g., dominance and minimax) are based on purely ordinal criteria whereas quantitative strategies (e.g., multi-attribute utility theory) are based on trade-offs or weighting of the dimensions. We proposed that the qualitative strategy of selecting the option that is superior on the more important dimension is more likely to be employed in the qualitative method of choice, whereas a quantitative strategy based on the trade-offs between the dimensions is more likely to be used in the quantitative method of matching. In this sense, the prominence effect may be attributable to the compatibility between the nature of the task and the nature of the strategy it invokes. For further discussion of strategy compatibility and its relationship to scale compatibility, see Fischer and Hawkins (1988).

Compatibility, like anchoring, can have a powerful effect on prediction and preference, yet people appear to have little or no conscious awareness of it, either inside or outside the laboratory. Such bias seems to operate at a very elementary level of information processing and it is doubtful whether it can be eliminated by careful instructions, or by monetary payoffs. Indeed, the use of incentives to promote careful responses has had little influence on the prevalence of preference reversals (Slovic & Lichtenstein, 1983).

The effects of compatibility described in this chapter represent a major source of violations of procedure invariance, namely the requirement that normatively equivalent elicitation procedures should yield the same ordering of options or events. The failure of procedure invariance complicates the task of the practitioner and the theorist alike. From a practical perspective, the present findings underscore the lability of judgments and choices, and make the elicitation task quite problematic. If the decision maker's response depends critically on the method of elicitation, which method should be used, and how can it be justified? At the very least, we need to use multiple procedures (e.g., choice, pricing, rating) and compare their results. If they are consistent, we may have some basis for trusting the judgment, if they are not, further analysis is required.

The assumption of procedure invariance plays an essential role in theories of rational choice. Behavioral research has also demonstrated consistent violations of description invariance by showing that different descriptions of the same decision problem can give rise to systematically different choices. Thus, alternative framings of the same options (e.g., in terms of gains vs. losses, or in terms of survival rates vs. mortality rates) produce predictable reversals of preference (Tversky & Kahneman, 1986). These failures of description invariance, induced by framing effects, and the failures of procedure invariance, induced by elicitation effects, represent deep and sweeping violations of classical rationality.

Attempts to describe and explain these failures of invariance require choice models of much greater complexity. To account for violations of description invariance, it seems necessary to introduce a framing process, including the determination of a reference point, which takes place prior to the valuation of prospects (Kahneman & Tversky, 1979). To account for violations of procedure invariance, it seems necessary to introduce multiple preference orders (obtained from choice, matching or pricing) and a contingent weighting model (Tversky et al., 1988) in which the tradeoff among attributes is contingent on the method of elicitation. These developments highlight the discrepancy between the normative and the descriptive approaches to decision making. Because invariance--unlike independence or even transitivity--is normatively unassailable and descriptively incorrect, it may not be possible to construct a theory of choice that is both normatively acceptable and descriptively adequate.

References

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. (1962). Some variables influencing the rate of gain of information. *Journal of Experimental Psychology, 63,* 105-110.

Delquie´, P., & de Neufville, R. (1988). *Response-modes and inconsistencies in preference assessments.* Unpublished manuscript, Massachusetts Institute of Technology.

Fischer, G. W., & Hawkins, S. A. (1988). *Preference reversals in multiattribute decision making: Scale compatibility, strategy compatibility, and the prominence effect.* Unpublished manuscript, Carnegie-Mellon University.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology, 46,* 199-210.

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

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

Hershey, J., & Schoemaker, P. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science, 31,* 1213-1231.

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

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

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

Schkade, D. A., & Johnson, E. J. (1988). *Cognitive processes in preference reversals.* Working paper, Department of Management, University of Texas.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Response mode, framing, and information-processing effects in risk assessment. In R. Hogarth (Ed.), *New directions for methodology of social and behavioral science: Question framing and response consistency,* no. 11 (pp. 21-36). San Francisco, CA: Jossey-Bass.

Slovic, P., & Lichtenstein, S. (1968). The relative importance of probabilities and payoffs in risk-taking. *Journal of Experimental Psychology Monograph Supplement, 78*(3), part 2 (b).

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavioral and Human Performance, 6,* 649-744.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review, 73,* 596-605.

Slovic, P., & MacPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance, 11,* 172-194.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327-352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Sci-*

ence, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business, 59* (4), Pt. 2, 251-278.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95* (3), 371-384.

Tversky, A., Slovic, P., & Kahneman, D. (1989). *The causes of preference reversal.* Unpublished manuscript, Stanford University.

Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus: Merrill.