ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 55, 78–94 (1993)

The Negative Effect of Probability Assessments on Decision Quality

IDO EREV

The Technion-Israel Institute of Technology, Haifa, Israel

GARY BORNSTEIN

The Hebrew University of Jerusalem, Jerusalem, Israel

AND

THOMAS S. WALLSTEN

The University of North Carolina at Chapel Hill

It is commonly assumed that uncertain information can be reduced to numerical probabilities without biasing preferences. It is also implicitly assumed in much research and many applications that people can express these probabilities. In contradiction to these assumptions Experiment 1 shows that the production of probability assessments biases decisions in an n-person game. Experiment 2 shows that the explicit assessment of numerical probabilities renders choices between gambles concerning future basketball events less optimal. These findings seem to be a result of overweighting the probability dimension relative to the payoff dimension given numerical judgments. Experiment 2 also suggests that without explicit numerical probability judgments subjects are less likely to violate the dominance principle. The theoretical and practical implications of the results are discussed. © 1993 Academic Press, Inc.

Most decision theorists assume that relevant information concerning the chances of a future event can be meaningfully and completely summarized by a unique probability value. For example, Savage (1954) used the term "quantitative personal probability" to refer to the numerical representation of probabilistic information in his "small world." He showed that such a representation, i.e., one satisfying the Kolmogoroff axioms, could be derived from a set of judgments provided they satisfy certain reasonable conditions. Other theorists ignored the complexity of

This research was supported by a grant from the Marshal foundation and by Grant BNS 8908554 from the U.S. National Science Foundation. We thank Lucia Kerpel and Chris Rodenbarry for their help in data collection. Address correspondence and reprint requests to Ido Erev, Faculty of Industrial Engineering and Management, Technion, Haifa, 32000, Israel.

0749-5978/93 $5.00

the relevant information altogether, and focused on decision making given information that is stated by numerical probabilities. For instance, Kahneman and Tversky (1979) proposed Prospect Theory to account for choices between numerically stated alternatives. In discussing the generalizability of Prospect Theory, Kahneman and Tversky wrote:

> The theory can also be extended to typical situations of choice, where the probabilities of outcomes are not explicitly given. In such situations, decision weights must be attached to particular events rather than to stated probabilities, but they are expected to exhibit the essential properties that were ascribed to the weighting function (p. 288).

Thus, even though Kahneman and Tversky (1979) present a theory inconsistent with Savage's assumptions, they, as well as many other theorists, seem to accept his view concerning the representation of uncertainty. In particular, they implicitly accept the assumption that uncertain information can be reduced to numerical probabilities without affecting preferences. We refer to this common assumption as the *Information Reduction (IR) assumption.*[1]

The IR assumption has also dominated behavioral decision research. As Kahneman, Slovic, and Tversky (1982) note in the preface to *Judgment under Uncertainty: Heuristic and Biases,* most of the research on decisions under uncertainty falls within either of two areas commonly labeled "judgment" and "decision making." Much judgment research (e.g., the work collected by Kahneman *et al.*, 1982) focuses on the way people form probability judgments, while much decision making research (e.g., the work that motivated Prospect Theory) focuses on choice behavior given objective probabilities, which are typically provided by the experimenter.

The IR assumption provides one justification for this historical division. If uncertain information is reduced to numerical judgments without affecting decisions, the two processes can be studied separately to improve

[1] The IR assumption can be formally stated as follows: For $i = 1,2$, let A_i = {outcome a_i if event E_i occurs; status quo otherwise} be two prospects, and A'_i = {outcome a_i with probability p_i; status quo with probability $1\text{-}p_i$} be the respective canonical prospects such that an individual is indifferent between A_i and A'_i ($A_i \sim A'_i$). Then

$$A_1 \text{ is preferred to } A_2 \ (A_1 > A_2)$$

if and only if

$$A'_1 \text{ is preferred to } A'_2 \ (A'_1 > A'_2).$$

The values of p_i for the canonical prospects can be found by basing the outcomes of A_i on the spin of a probability wheel and systematically manipulating p between 0 and 1 until a value of p_i is found such that $A_i \sim A'_i$. We refer to this value as the numerical representation of the information the individual possesses concerning the likelihood of event E_i.

experimental control. On the other hand, if the IR assumption does not hold the division can lead to misleading conclusions.

Given the wide implicit acceptance of the IR assumption it is not surprising that it serves as a basis for decision analysis. To derive the practical recommendations it is a natural step to assume further that an individual can actually state a numerical value that expresses his or her information concerning the likelihood of relevant events. For example, it is often recommended that experts use probability numbers to communicate information about likelihoods of future events (Behm & Vaupel, 1982; von Winterfeldt & Edwards, 1986). In particular, numerical judgments (e.g., 80% *chance*) are supposed to lead to better decisions than do verbal assessments (e.g., *likely*).

Recent experimental work (Budescu & Wallsten, 1990; Erev & Cohen, 1990; Gonzalez-Vallejo, Erev, & Wallsten, in press), however, has failed to support this expectation. Numerical communication of uncertain information did not lead to more efficient decisions than did verbal communication of the same information, and under certain conditions, it actually led to worse decisions. For example, in one of these experiments (Gonzalez-Vallejo *et al.*, in press) subjects were run in pairs, with one member randomly chosen to be the decision maker (DM) and the other to be the performer. The performer had to win points in simple tasks (such as throwing a chip to a certain target), and the decision maker had to choose among gambles such as

$x if the performer's score in task Y is at least 2 points;
$0 otherwise.

The choice task (similar to that used in Experiment 2, below) was constructed so that a preference ordering was derived over a set of six gambles. The decision maker acted after the performer communicated his or her judgments of the chances of achieving the various scores. The judgments were expressed numerically for some choices and verbally for others. All earnings were divided between the two subjects at the end of the experiment.

It turned out that the preference ordering correlated relatively less strongly with the payoff dimension when uncertainty was communicated numerically than verbally. For some sets of gambles, this tendency was consistent with expected value (EV) considerations, and consequently subjects earned more money when judgments were expressed numerically than verbally. However, in other sets of gambles this tendency was inconsistent with the EV rule, and in these cases verbal judgments led to significantly higher profits than did numerical judgments. Thus, it appeared that the form of the uncertainty representation influenced the relative weight accorded to the two dimensions of the gambles, and this effect was not always beneficial.

These unexpected results raise the possibility that other widely held beliefs derived from the IR assumption may also be false under certain conditions. The present paper focuses on two such beliefs. The first, which we refer to as the *maximization hypothesis*, is that decision makers are more likely to make decisions consistent with their preferences by explicitly producing prior numerical assessments of the probabilities of relevant future events. Hence, if they seek to maximize a certain function, probability assessments will aid in achieving this goal. This argument was recently stated by von Winterfeldt and Edwards (1986) who wrote

> . . . quantification of subjective variables—probabilities and values—can in itself be very useful, even if further analysis is left to non-decision-analytic approach or to the intuition of the decision maker (p. 23).

Under the IR assumption it is easy to justify this argument. Since no information is lost when knowledge is reduced to a single number, a parsimonious representation of the decision problem can only help to eliminate errors (i.e., eliminate decisions that are inconsistent with the DMs' preferences).

The second belief, which we refer to as the *independence hypothesis*, posits that decisions are independent of (not biased by) the production of numerical probability judgments. The IR assumption suggests that subjects should make the same decisions regardless of whether they rely on their personal information or on numerical representations of this information. Therefore, asking subjects to state their probabilities should not change their preferences. Indeed, many researchers interested in the relationship between subjects' personal knowledge and choice behavior ask subjects to assess probabilities of relevant future events, on the assumption that the production of these judgments does not systematically affect subsequent behavior.

Some of our own previous work in the field of *n*-person games (e.g., Bornstein & Rapoport, 1988; Erev & Rapoport, 1990; Rapoport, Bornstein, & Erev, 1989), as well as work by others in this field (e.g., Dawes, McTavish, & Shaklee, 1977; Rapoport & Eshed-Levi, 1989), is based at least in part on the independence claim. In this research, subjects (players) typically were asked to assess the probability that other players will make certain decisions and then were asked to make their own decision. Experiment 1 was designed to study the independence hypothesis in the context of experimental games by varying the order of assessments and decisions. It utilizes a simple version of an *n*-person game called the *intergroup public goods* (IPG) game recently devised by Rapoport and Bornstein (1987).

EXPERIMENT 1

Method

Subjects (160 undergraduate students at the Hebrew University of Jerusalem) were run in groups of 10. Upon arrival at the laboratory, the subjects were seated in a single room with arrangements to ensure their privacy and were assigned to one of two subgroups consisting of five members each.

Each subject received a promissory note of 5 Israeli shekels (about \$3.50) and was told that he or she could either keep the note and cash it in at the end of the session or invest it. Subjects were instructed that members of the subgroup with the higher number of investors would each receive a monetary reward of r units (r = 15 shekels for half of the subgroups and 25 shekels for the other half) and that in case of a tie (equal number of investors in each group) members of both subgroups would each receive a reduced reward of $r/2$ units (7.5 or 12.5 shekels).

At that stage, half of the subjects (the assessment-first condition) responded to a questionnaire in which they were asked to assess:

1. the probability that each of the other members of their subgroup would invest;
2. the probability that each of the five members of the other subgroup would invest;
3. the probability that their own decision would affect the outcome of the game;
4. the probabilities of all possible combinations of numbers of investors in the two subgroups.

After completing the questionnaire, these subjects were each given a form on which they indicated their decision whether to invest. The other half of the subjects (the decision-first condition) were given the decision form first, and answered the questionnaire subsequently.

The size of the reward and the other condition were balanced over subgroups.

Note that the game-theoretical equilibrium solution for this game is to invest in all conditions. However, as we have shown elsewhere (Bornstein, Erev, & Rosen, 1990), people do not always follow this Pareto deficient prediction. Hence, it is not clear what is the best strategy for a profit-maximizing subject.

Results and Discussion

Table 1 presents the proportion of investors by r (the reward size) and of the order in which decisions and assessments were made. An analysis of variance revealed a significant interaction ($\chi^2[1] = 5.79$, $p < .02$)

TABLE 1
PROPORTION OF INVESTMENT BY CONDITION AND REWARD SIZE

Reward size	Condition: Decision-first	Assessment-first	Mean
15	.475	.650	.562
25	.725	.525	.625
Mean	.600	.587	.594

between the reward size and the order conditions. When subjects made their decisions before judging the probabilities of the different results, the reward size affected the decisions—72.5% of the subjects chose to invest when the reward was 25 shekels, and only 47.5% did so when the reward was 15 shekels ($p < .025$). On the other hand, when subjects were asked to judge the probabilities first, the reward size had no effect on their decisions ($p > .05$). That is, the reward size effect was eliminated by assessing probabilities before making decisions. This result is inconsistent with the independence hypothesis, which predicts that the production of the assessments will not change the subject's preferences.

Moreover, it should be noted that the decisions following probability assessment were not more optimal than those preceding it (based on either a game-theoretic or an EV criterion of optimality). As can be seen in Table 1, the proportion of subjects deciding to invest (in accordance to the game-theoretic solution) was lower following the probability assessments. Although an EV strategy cannot be defined in the absence of knowing the probability that one's decision to invest will be essential in obtaining the reward, investment is more likely to be optimal when $r = 25$ than when $r = 15$. Thus, under this criterion too, the assessments did not improve the decisions.

The violation of the independence assumption, just as the violation of the communication mode hypothesis, can be explained by the argument that numerical probability judgments cause DMs to pay less attention to the possible payoffs in their choices. Consequently, the decisions of subjects in the assessment-first condition were not affected by the reward size. It seems that, consistent with suggestions of Tversky, Sattath, and Slovic (1988), DMs are likely to choose according to the probability dimension when numerical judgments are available; however, they are less likely to do so in the absence of numerical judgments.

Experiment 2 was designed to examine whether the same phenomenon can be observed when the DM's profit depends only on his or her own knowledge and choices. More specifically, Experiment 2 was designed to test whether explicit assessment of relevant numerical probabilities, by the DM him- or herself, decreases the effect of the possible outcomes on subsequent decisions. If this indeed happens, subjects may earn less money when requested to assess probabilities, in violation of the maximization hypothesis (assuming that the subjects are trying to maximize profit). For a weaker test of the maximization hypothesis Experiment 2 was also designed to compare the likelihood of violations of the dominance principle when choices are made following explicit probability judgments and when choices are made in the absence of such assessments.

EXPERIMENT 2

Method

Subjects (61 undergraduate students at the University of North Carolina at Chapel Hill) were run in groups of about 10. Their task was to choose among (rank) gambles in three different sets, where each gamble promised a certain amount of money if a particular event occurred in one of the nine remaining UNC–CH basketball games. (A list of these nine games was provided to the subjects.) The subjects were told that at the end of the experiment one of the games would be randomly chosen to be their "payoff game" and one of the three gambling sets would be randomly selected to be their "payoff set." It was emphasized that they would be paid in accordance with the results of the payoff game and their choices in the payoff set, and that their goal was to maximize earnings.

Table 2 presents one of the three gambling sets (with the choices of a hypothetical subject). Ignore temporarily the columns headed by *P* and (EV), which were not present on the sheets actually given to the subjects. Each gamble is written as the base outcome (labeled "value" in the table) to be won if the corresponding event occurred in the payoff game. The subjects made choices about the gambles in the seven columns opposite them. In the first column, they chose only one gamble, and, therefore, had to select the best one. In the second column, subjects chose two gambles that they believed to be the best, and so forth. Thus, in the seventh column, they chose seven of the gambles (all of the gambles but the worst one). The choices were made by marking X's in the columns.

The subjects were instructed that after their payoff game was played, they would win on all of the gambles in the selected payoff set involving events that actually occurred. Their earnings would be determined by multiplying the base outcome of each successful gamble by the number of times it was selected.

For example, assume that the selected payoff set is the one presented in Table 2, and that the events in gambles B, C, E, and H occurred in the payoff game. The hypothetical subject who made the decisions in Table 2

TABLE 2

ONE OF THE THREE GAMBLING SETS FROM EXPERIMENT 2 WITH HYPOTHETICAL SUBJECT'S CHOICES, ASSESSMENTS, AND E

Gamble	You win if the following occurs in your payoff game	Value	*P*	(EV)	Choose the best gamble	Choose the best 2 gambles	Choose the best 3 gambles	Choose the best 4 gambles	Choose the best 5 gambles	C b ga
A	J. Lebo will score more than D. May.	$.15	80	(12)		X	X	X	X	
B	J. R. Reid will take more rebounds than J. Denny.	$.15	100	(15)	X	X	X	X	X	
C	S. Williams will take more rebounds than P. Chilcutt.	$.30	60	(18)				X	X	
D	K. Madden will score more than H. Davis.	$.10	80	(8)						
E	R. Fox will score more than S. Williams.	$.20	60	(12)						
F	P. Chilcutt will take at least as many rebounds as S. Williams	$.30	40	(12)						
G	K. Madden will pass more assists than K. Rice.	$.70	30	(21)			X	X	X	
H	P. Chilcutt will score more than 15 points.	$.60	15	(9)					X	

would earn \$.15 seven times for gamble B, \$.30 four times for gamble C, \$.20 two times for gamble E, and \$.60 three times for gamble H. Thus, the earnings in this case would be:

$$(\$.15)7 + (\$.30)4 + (\$.20)2 + (\$.60)3 = \$4.45.$$

Note that from the subjects' point of view the choice task was equivalent to a ranking task. To maximize expected profit they had to choose the highest EV gamble in the first column, the two highest EV gambles in the second column, and so on. As in previous work which used this ranking-choice procedure (Erev & Cohen, 1990; Erev *et al.*, 1989) the subjects did not find the task difficult to understand.

The gambling sets were designed to achieve three goals. The main goal was to study whether the act of estimating probabilities affects the relative weights the DMs give to the probabilities and the outcomes in choosing among the gambles. The second goal was to look at the effect of probability assessment on the amount of money the DMs earned. The third goal was to detect possible violations of the dominance principle. In one of the gambling sets (the one presented in Table 2) the choice of each of the four pairs of gambles, A&B, A&D, B&D, and D&E promised at most \$.30 or \$.25 each, whereas the combined choice of gambles C and F promised \$.30 for sure (since event F is the complement of event C). Clearly, each of the four pairs presented above is dominated by the C&F pair. Hence, a choice of any one of these pairs over the C&F pair is a violation of the dominance rule. For example, if one chooses gambles A and B over C and F (as illustrated in the second choice column of Table 2) one is violating the dominance principle. Following Tversky and Kahneman (1981) we conceptualize violations of this type as "editing errors," assuming that subjects will fail to make the dominance choice only if they do not edit the C and F choices as \$.30 for sure.

Another gambling set was designed to elicit violation of the dominance rule through the conjunction fallacy (Tversky & Kahneman, 1982). This set included the following two gambles:

A—\$.25 if R. Fox will score more than 22 points;

B—\$.25 if R. Fox will score more than 22 points and UNC will win.

Because the event in gamble A includes the one in gamble B, a choice of gamble B over gamble A is a violation of the dominance rule.

Finally, the subjects were asked to answer 10 multiple choice questions in a "UNC Basketball Quiz," designed to measure their level of expertise. The quiz included questions of the type:

The best scorer in the first UNC–Duke game this season was:

a J. R. Reid b. K. Madden c. K. Rice d. S. Williams.

Just as in Experiment 1, there were two order conditions, assessments-first and decisions-first. In the assessment-first condition (n = 32) the subjects first assessed the chances of winning each of the 24 gambles by writing a number from 0 to 100 in the empty column next to each gamble. (This is the major column that has the P and EV entries in Table 2. Note that the EV column was never present for the subjects.) They then answered the quiz and indicated their choices within the different gambling sets. In the decision-first condition ($n = 29$) subjects first answered the quiz, then made their decisions, and finally assessed the chances of winning the gambles.

Results and Discussion

The probability judgments. Recall that the subjects assessed the chances of winning each of the 24 gambles based on the events that occurred in one game randomly chosen from the remaining nine games. In order to evaluate the accuracy of the subjects' probability judgments we calculated the correlation between the judgments and the proportions of the nine games in which the events actually occurred. These post hoc relative frequencies are the best assessment of the "real" probabilities.

The average correlation between the subjects' judgments and the post hoc relative frequencies was 0.53 in the decisions-first condition ($p < .0001$) and 0.51 in the assessments-first condition ($p < .0001$). These average correlations are quite impressive given the fact that the correlation was only 0.70 between the relative frequency estimates of the events based on the 15 games preceding the experiment and those based on the 9 subsequent games.

Analysis of covariance was used to test whether experimental condition and expertise (as measured by the quiz score) affected subject accuracy (i.e., the correlation between the judgments and the post hoc relative frequencies). Whereas expertise was found to be a good predictor of accuracy ($F[1,57] = 16.44$, $p < .0002$), the experimental condition was not ($p > .05$). These results suggest that the subjects' judgments depended on their personal knowledge, but not on the experimental condition (at least according to our accuracy criterion). It seems that subjects did not change their assessments systematically as a result of prior decision making.

The nature of the choices. The column P in Table 2 shows the (hypothetical) subject's probability judgments. To facilitate the explanation of the analysis presented below we added the column (EV) to the table, showing subjective expected values (which were calculated by multiplying the value by the subject's probability assessment.)

To study the weights given to the dimensions of the gambles in the ranking-choice task we used the V statistic (Nelson, 1984), defined as the proportion of pairwise agreements between two rankings of a set. Consider two measures (X and Y), each of which induces a rank order of n

stimuli (n gambles in our case). The n stimuli yield $n(n-1)/2$ pairs, each of which may be ordered in the same or opposite direction by X and Y. Excluding pairs to which either X or Y gives a tie rank, $V[X,Y]$ is the proportion of the pairs in which the stimulus that is ranked higher by X is also ranked higher by dimension Y. That is,

$$V[X,Y] = \frac{\text{No. agreements}}{\text{No. agreements} + \text{No. disagreements}}.$$

Note that V is a simple linear transformation of Goodman and Kruskal's (1954) gamma (G) measure of association ($V = .5G + .5$). We prefer to use this measure over similar and more traditional measures of similarity because of its comparability to choice proportion, a common statistic in decision making research. Conveniently, the interpretation of V as a proportion makes linear manipulation of this measure meaningful.

Our main dependent variable is the V score comparing the ranking of a set of gambles by the decisions (D) and by the payoffs ($). Note that V[D,$] is the proportion of pairs in which the gamble providing the higher possible outcome is also higher in the preference ordering. Thus, we refer to V[D,$] as a measure of $-preferences. Additional dependent variables are V[D,P] which measures P-preferences, where P represents the judged probabilities of the events, and V[D,EV], which measures EV-preferences, with EV calculated as in Table 2. It is important to point out that for a given pair of gambles the EV-ordering may coincide with the $-ordering or with the P-ordering, and that in some cases the three orderings may all agree, but over all pairs of gambles the three orderings will be distinct.

Table 3 presents the mean preference measures of the different types over subjects in the two experimental conditions. The difference between the two groups in the average $-preference measure is highly significant ($t[59] = 3.50$, $p < .001$). This difference is consistent with the findings of Erev *et al.* (1989), and with the results in Experiment 1; in all cases the effect of the payoffs on the decisions was decreased in the presence of

TABLE 3
AVERAGE NUMBER OF $-PREFERENCES, P-PREFERENCES, AND EV-PREFERENCES IN THE TWO CONDITIONS

	Condition	
	Decision-first	Assessment-first
$-preferences (V[D,$])	.44	.34
P-preferences (V[D,P])	.83	.92
EV-preferences (V[D,EV])	.72	.59

numerical probability judgments. The robustness of this phenomenon suggests that different decisions are made in the presence of numerical probability judgments than in their absence.

Before discussing the *P*-preferences and EV-preferences, it is important to consider two difficulties with the interpretation of these measures. First, recall that they are based on the subjects' expressed probability judgments as well as on their choices, unlike the \$-preferences measure, which is affected by the subjects' choices alone. The second difficulty arises from the fact that the *P*- and the EV-preferences are not independent of the \$-preference. Nevertheless, these measures are of interest since they slow the relationship between the probability judgments and the choices. To the extent that the probabilities expressed by the subjects represented their true knowledge (suggested by the finding that accuracy depended on expertise, but not on the experimental condition) we can learn about the relation between their knowledge and their decisions in the two experimental conditions.

Three subjects in the decisions-first condition did not complete the probability judgments, and were excluded from the *P*-choices and EV-choices analyses presented in Table 3. Table 3 shows that the remaining 26 subjects in the decisions-first condition demonstrated significantly ($t[56] = 2.99$, $p < .005$) fewer *P*-preferences than did the subjects in the assessments-first condition. Also, subjects in the decisions-first condition indicated significantly more EV-preferences than did subjects in the assessments-first condition ($t[56] = 3.86$, $p < .0005$). Strictly speaking, these two plus the previous t test are not independent of each other, but the important message is in the magnitudes of the effects, not in their statistical significance levels.

In summary, the analyses show that the request to assess the relevant probabilities decreased the likelihood of EV-preferences. Instead, it seems that the subjects in the assessments-first condition were more likely to rely heavily on their probability judgments.

Profit analysis. After the nine basketball games had been played, the hypothetical profit of each decision maker was calculated as the total amount of money he or she would have earned if all three sets of gambles had actually been played in all nine games. (Recall that the subjects were actually paid according to only one of these $3 \times 9 = 27$ combinations. However, since they did not know when making the decisions which one it would be, they were motivated to choose the best gambles in each set considering all the games.) In order to test the validity of the hypothetical profit as a measure of the quality of the decisions, it was compared to the expected profit from random choices. Since there were 28 choices per set, each gamble would be selected at random $28/8 = 3.5$ times. The expected profit of such random choices over the three sets is \$97, which is significantly ($t[60] = 5.95, p < .0001$) lower than the $107 hypothetical profit of the average subject. The expected profit measure was also found to be positively correlated ($r = .21$) with the quiz score. The correlation is marginally significant ($t[60] = 1.68, p < .1$). These results suggest that the hypothetical profit variable is sensitive to the quality of the decisions.

The mean hypothetical profit of the decision-first subjects was $111 and that of the assessment-first subjects was $104. The difference between the two conditions is significant ($t[59] = 2.1, p < .04$). That is, in violation of the maximization hypothesis, assuming subjects prefer to make more money than less, subjects who made explicit assessments of the relevant probabilities before making their decisions earned smaller (hypothetical) profits than did subjects who made their decisions first.

The dominance rule. Each subject could exhibit five violations of the dominance rule (up to four "editing errors" and one conjunction fallacy). The mean number of editing errors was 1.24 in the decision-first condition and 2.0 in the assessment-first condition. The difference is significant ($t[59] = 2.26, p < .03$).

The conjunction fallacy was made by 31% of the subjects who made their decisions first, and by 41% of the subjects who estimated probabilities first. Although the violations of the conjunction rule are more frequent in the assessments-first condition, the difference is nonsignificant ($p > .05$). A large difference between the two experimental conditions was found in the number of subjects who did not exhibit any violations of the rule. Whereas 11 of the 29 decisions-first subjects (39%) did not violate the dominance principle, only 3 of the 32 assessments-first subjects (9%) avoided errors of this type.

These results are consistent with the EV-choices and the profit analyses presented above. All of these analyses suggest that the request to express probability judgments rendered the decisions of poorer quality.

GENERAL DISCUSSION

The two experiments presented above show that after estimating the probabilities of relevant future events decision makers tended to pay less attention to payoffs when choosing between risky alternatives. These results contradict the independence and the maximization hypotheses derived from the IR assumption in conjunction with the assumption that individuals can express their personal information numerically. Experiment 1 demonstrates that asking subjects in an n-person experimental game to assess the probabilities of the behavior of other players affects their choices. Experiment 2 shows that quantification of subjective probabilities can lead to a decrease in profits and an increase in violations of the dominance rule. In a similar manner, research by Gonzalez-Vallejo *et al.* (in press) has shown that under certain conditions verbal communication of uncertain information leads to higher profits than does numerical communication.

Surprisingly behavior in Experiment 2 was more consistent with expected utility (EU) theory in the absence of quantified probability judgments than in their presence. Subjects in the decisions-first condition were less likely to violate the dominance principle and showed more balanced weighting of probability and outcomes. Thus, whereas this paper suggests a deleterious effect of explicitly stating numerical probabilities, it also suggests that Savage's axioms may be a better approximation of human behavior when probabilities are not assessed than when they are.

This pattern of results stands in sharp contradiction to the early criticism of the EU model. Recall that it has been argued (e.g., Simon, 1955) that the effort needed to maximize expected utility is too high for such a model to be descriptive. Consistent with this line of thinking, most recent critics of the EU model have limited their research to problems stated in numerical form on the assumption that if the model is violated under these conditions, it is definitely not useful for problems that are less well structured. The results of Experiment 2 suggest that this basic assumption may be incorrect.

Toward a Theoretical Understanding

Experimental results in behavioral decision research are heavily affected by the experience the subjects bring to the laboratory (Thaler, 1987; Tversky, 1981). Decision researchers do not argue that people cannot learn to behave in a way that benefits them, but rather that real life experience does not always lead to the behavior described by the EU theory. One of the main goals of behavioral decision research is to learn how people decide in light of their personal experience.

Our results, which may seem surprising when DMs are thought of as limited capacity sequential information processors, are actually quite reasonable when the DMs' experience is taken into account. Obviously, we all have more experience choosing among prospects when the probabilities of the outcomes are not explicitly stated, than among numerically stated gambles. As von Winterfeldt and Edwards (1986) pointed out, people make choices hundreds of times a day between prospects with outcomes that depend on future events. On the other hand, situations in which people must decide between numerically stated alternatives are rare. It can be assumed that their vast experience led our subjects toward reasonable, although not perfect, decisions in the absence of numerical probabilities. Their unfamiliarity with decision making given numerical expressions of uncertainty led them toward overweighting the "new" representation of uncertain information.

An extreme analogy may help to make our point. Consider professional golfers who play as if they combine information concerning distance and direction of the target, the weight of the ball, and the speed and direction of the wind. Now assume that we ask them to play in an artificial setting in which all the information they naturally combine in the field is reduced to numbers. It seems safe to say that the numerical representation of the information will not improve the golfer's performance. The more similar are the artificial conditions we create to the conditions with which the golfers are familiar, the better will be their performance. One can assume that decision making expertise, like golf expertise, is improved by experience, but not always easily generalized to new conditions.

A similar perspective has been offered by Hammond, Hamm, Grassia, and Pearson (1987) based on Brunswik's (1957) approach to studying cognition. They propose a cognitive continuum from analytic (involving task-specific strategies) to intuitive. The ideal point to operate on the continuum depends on task properties and on how the information is represented. Performance can be impaired if the task demands and information representation conflict.

Two issues concerning the generalizability of our finding should be discussed. The first concerns the task features that created the new, unfamiliar situation in the assessments-first condition. The second question concerns the subjects' experiences that they may have brought to the laboratory.

Task features. A few aspects of the assessments-first condition may have affected the decisions: First, it seems reasonable that by asking the subjects to judge probabilities we signaled to them that they are important. As a result, the probabilities were overweighted and/or their scale values were changed. Moreover it might be argued that the decisions were biased by the manner in which we requested the assessments. Second, it is possible that the DMs were attempting to be consistent with their assessments. Third, the existence of the assessments may have created an illusion of transforming the task from decision making under uncertainty to decision making under risk.

Possibly, all claims are correct, and under other circumstances DMs can assess probabilities without biasing their subsequent choices. It remains for future research to compare the biasing effects of various elicitation procedures.

The critical experience. As in much behavioral decision research our subjects were undergraduate students. It is often argued that the consistencies observed in undergraduate students' decisions may not describe choice behavior of experts or of more experienced decision makers. It must be pointed out in this regard that our subjects in Experiment 2 were very knowledgeable in the subject matter underlying the gambles. What they were not experienced in, most likely, was selecting among gambles, especially not after judging numerical probabilities. Probably with additional experience in choosing among numerically stated gambles, the results would change and the bias would be eliminated. This latter point is important, in that most decision makers in the real world have little experience in choosing among numerically defined alternatives.

It should be pointed out that knowledge of decision theory, per se, may not prevent violations of the IR assumption. Recall that Savage made a mistake when first faced with the Allais paradox. In the version of the paradox that led to his error, uncertain information was presented with numerical probabilities. To convince himself that he had made a mistake, Savage (1954, p. 103) reformulated the problem by substituting events (drawing of lottery tickets) for the probabilities. Although this anecdote is only suggestive, it does seem that Savage's intuition, like our subject's preferences, was more consistent with EU theory when uncertainty was not presented numerically.

REFERENCES

Behm, R. D., & Vaupel, J. W. (1982). *Quick analysis for busy decision makers*. New York: Basic Books.

Bornstein, G., & Rapoport, A. (1988). Intergroup competition for the provision of step-level public goods: Effects of preplay communication. *European Journal of Social Psychology*, **18**, 125–142.

Bornstein, G., Erev, I., & Rosen, O. (1990). Intergroup competition as a structural solution for social dilemmas. *Social Behavior*, **5**, 247–260.

Brunswik, E. (1957). Scope and aspects of the cognitive problem. In H. Gruber, K. Hammond, & B. R. Jessor (Eds.), *Cognition: The Colorado Symposium*. Cambridge, MA: Harvard Univ. Press.

Budescu, D. V., & Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. *Organizational Behavior and Human Decision Processes*, **46**, 240–263.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma game. *Journal of Personality and Social Psychology*, **35**, 1–11.

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes*, **45**, 1–18.

Erev, I., & Rapoport, A. (1990). Provision of step-level public goods: The sequential contribution mechanism. *Journal of Conflict Resolution*, **34**, 401–425.

Gonzalez-Vallejo, C., Erev, I. & Wallstens T. S. (in press). Do decision quality and preference order depend on whether probabilities are verbal or numerical? *The American Journal of Psychology*.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, **49**, 732–764.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, **17**, 753–770.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge Univ. Press.

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

Nelson, T. O. (1984). A comparison of current measures of accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109–133.

Rapoport, A., & Bornstein, G. (1987). Intergroup competition for the provision of binary public goods. *Psychological Review*, **94**, 291–299.

Rapoport, A., & Bornstein, G. (1989). Solving public good problems in competition between equal and unequal size groups. *Journal of Conflict Resolution*, **33**, 460–479.

Rapoport, A., Bornstein, G., & Erev, I. (1989). Intergroup competition for public goods: Effect of unequal resources and relative group size. *Journal of Personality and Social Psychology*, **56**, 748–756.

Rapoport, A., & Eshed-Levy, D. (1989). Provision of step-level public goods: Effects of greed and fear of being gypped. *Organizational Behavior and Human Decision Processes*, **44**, 325–344.

Savage, L. J. (1954). *The foundation of statistics*. New York: Wiley.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, **69**, 99–118.

Thaler, R. (1987). The psychology of choice and the assumption of economics. In A. E. Roth (Ed.), *Laboratory experimentation in economics: six points of view*. Cambridge: Cambridge Univ. Press.

Tversky, A. (1981). L. J. Cohen, again: On the evaluation of inductive intuition. (Open Peer Commentary to L. J. Cohen (1981). Can Human irrationality be experimentally demonstrated?) *Behavioral and Brain Sciences*, **4**, 354–356.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representatives. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristic and Biases* (pp. 84–100). New York: Cambridge Univ. Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, **211**, 453–458.

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

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

RECEIVED: September 27, 1990