Organizational Behavior and Human Decision Processes
Vol. 83, No. 2, November, pp. 260–281, 2000
doi:10.1006/obhd.2000.2909, available online at http://www.idealibrary.com on IDEAL®

Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation

Ilan Yaniv and Eli Kleinberger

Hebrew University of Jerusalem, Israel

Our framework for understanding advice-taking in decision making rests on two theoretical concepts that motivate the studies and serve to explain the findings. The first is egocentric discounting of others' opinions and the second is reputation formation for advisors. Advice discounting is attributed to differential information, namely, the notion that decision makers have privileged access to their internal reasons for holding their own opinion, but not to the advisors' internal reasons. Reputation formation is related to the negativity effect in impression formation and to the trust asymmetry principle. In three studies we measured decision makers' weighting policy for advice and, in a fourth study, their willingness to pay for it. Briefly, we found that advice is discounted relative to one's own opinion, while advisors' reputations are rapidly formed and asymmetrically revised. The asymmetry implies that it may be easier for advisors to lose a good reputation than to gain one. The cognitive and social origins of these phenomena are considered. © 2000 Academic Press

Taking advice is a central component of realistic decision making, since decision problems generally do not come as complete textbook problems with the relevant information as givens. Instead, decision makers engage in interactive processes in order to build the relevant informational basis. In particular, they solicit the opinions of worthy advisors, assess the merit of each opinion, and combine them. Examples include asking a fellow bus passenger about a street address, consulting a colleague's opinion for a research project, and genetic counseling regarding reproductive decisions. A variety of social and cognitive

This research was supported by grants from the Israel Foundations Trustees and the Israel Science Foundation to Ilan Yaniv. The support of the Rotman Center at the Hebrew University is also acknowledged. Portions of the research are included in a Master's thesis submitted by Eli Kleinberger to the Department of Psychology.

Address correspondence and reprint requests to Ilan Yaniv, Department of Psychology, Hebrew University, Jerusalem, Israel 91905. E-mail: ilan.yaniv@mscc.huji.ac.il.

0749-5978/00 $35.00

processes may thus play a role even in individual decision making, including asking for other people's advice, evaluating their expertise, combining their opinions, and attempting to reconcile inconsistencies (Harvey & Fischer, 1997; Sniezek & Buckley, 1995).

Receiving advice often exposes the decision maker to a potential conflict between his or her own initial opinion and the advice, and hence to a difficulty in combining the opinions. Consider a book editor who feels that a particular chapter would make an excellent contribution to a certain book. The editor then receives a lukewarm review of that chapter. How might the editor revise his or her opinion? The key question in many practical situations is to decide just how much weight one ought to place on a particular piece of advice so as to achieve an efficient combination of the opinions. In particular, a decision maker's weighting policy might entail completely ignoring the other opinion, some adjustment of one's own opinion toward the other, or complete adherence to the other opinion. The process by which decision makers form and revise such weighting policies is the topic of the present research.

In this article, we investigate how decision makers integrate advice and, in particular, how they revise their own opinions on the basis of another person's opinion that differs from their own. We suggest a preliminary framework for understanding decision makers' use of advice. Two theoretical concepts motivate the studies and serve to explain the findings. The first is egocentric discounting of other's opinion and the second is reputation formation for advisors.

We start by briefly reviewing normative and descriptive views on combining opinions and using advice. We report four studies. In Studies 1–3 we analyze decision makers' weighting policies and in Study 4 we consider decision makers' willingness to buy estimates from an advisor whose advice they had been given on earlier trials. The results show evidence for discounting as well as a fast learning process such that advisors acquire a reputation almost instantaneously. These findings are related to social and cognitive aspects of interactive decision making.

Combining Judgments

We take the view that a subjective estimate of the truth is the outcome of an error-prone mental process. Thus a subjective estimate can be expressed as the sum of the true answer and error components, including systematic bias and random noise (Einhorn, Hogarth, & Klempner, 1977; Wallsten, Budescu, Erev, & Diederich, 1997). The implication of this statistical view is that the aggregation of a number of opinions (drawn randomly from a population) improves accuracy. The effect of aggregation accuracy has indeed been documented in the judgment and forecasting literature across a wide range of domains, ranging from perceptual assessment tasks (Zajonc, 1962; Harvey & Fischer, 1997; Hill, 1982; Yaniv & Hogarth, 1993) to medical diagnoses (Winkler & Poses, 1993) and business–economic forecasts (Ashton & Ashton, 1985; Graham, 1996; Libby & Blashfield, 1978; Zarnowitz, 1984). (It has also been noted anecdotally in a chess-playing situation by Seife [1996].)

A common normative approach to the combination problem essentially treats estimates as data to be summarized in terms of measures of central tendency and dispersion. A recommendation for combining opinions is that, other things being equal, opinions in a sample should be weighted equally. As the research cited above suggests, there are strong empirical and theoretical (normative) justifications for combining estimates from different sources and, in particular, for equal weighting (simple averaging). Differential weighting of opinions may be warranted only if additional information about the forecasters is available, such as expertise, past accuracy, and confidence (Ashton & Ashton, 1985; Sniezek & Henry, 1989; Yaniv, 1997). Indeed, simple averaging has been the most frequent combination scheme in the studies cited above, so it will be used as a reference in discussing the weighting policies that respondents use (Dawes, 1979; Einhorn & Hogarth, 1975; Harvey & Fischer, 1997).

Discounting Advice and Reputation Formation

From an external, objective point of view, a decision maker's opinion and the advice received are data and as such, other things being equal, they should be equally weighted. Consider, however, the decision maker's perspective and the psychological forces that shape the weighting of his or her own opinion and the advice. The decision maker may not consider both opinions to be on an equal footing. Individuals normally have privileged access to the reasons that lead them to hold their own opinion, but only limited access (if any) to the reasons that lead other people to hold their opinions. This fundamental disparity between the access to the support for one's own opinion and the access to the support for another person's opinion sets the stage for differential weighting of own opinion versus the advice. It is commonly accepted in behavioral analysis of judgment that the psychological weight assigned to a hypothesis or an estimate is directly related to the amount of support that can be retrieved for that estimate (e.g., Tversky & Koehler, 1994). This line of reasoning leads to the discounting hypothesis. Namely, when faced with a discrepancy between their own opinion and some advice, decision makers tend to discount the advisor's opinion in the combination process.

How general or pervasive is discounting expected to be? Is it immune to feedback? How is it affected by the history of the relationship between the decision maker and the advisor—in particular, the advisor's previous successes and failures? We suggest that decision makers form a reputation for the advisor which mediates their weighting policy. Reputation formation may be described as the process of forming an impression about an advisor's quality from previous experience. Our research simulates an interactive situation whereby the decision maker receives feedback online and thus gains experience with the advisor. We compare the weighting policies created for a competent advisor (who is substantially better than the average respondent), a poor advisor (who is worse than the average respondent), and an average advisor.

How does reputation evolve? The term *reputation* is used here to describe what a decision maker learns about an advisor through experience. Experience

may consist of a series of trials in which the decision maker obtained advice from the advisor. There are reasons to expect that good and poor advice might have a differential impact on the evolution of reputation. One basis for this conjecture comes from the social psychology of impression formation. In a common research paradigm, lists of trait adjectives (ascribed to an actor) or lists of behaviors are used to elicit impression ratings experimentally. Studies have shown that impression judgments are inordinately influenced by an actor's more negative attributes (Fiske, 1980; Hamilton & Huffman, 1971). For instance, in judging an individual's morality, unfavorable information (e.g., an instance of dishonest behavior) tends to be more influential than favorable information (an instance of honest behavior). The term *negativity bias* reflects the finding that negative information causes impression ratings to deviate from the predictions of an equal-weight averaging model. According to one recent account of the negativity bias, negative information is perceived as more diagnostic of an actor's true character than positive information (Skowronski & Carlston, 1989). One major reason is that favorable information is consonant with our social norms and hence conveys less information about an individual's personality. In contrast, unfavorable information indicates a manner in which the person deviates from social norms and indicates something unique about him or her as an individual.

Within the judgment and decision-making area, a similar idea is echoed in Slovic's notion of trust asymmetry (1993). Slovic notes a pattern of asymmetry in people's expressions of trust in the advice of experts regarding environmental and technological risks (such as radiation or nuclear waste sites). He suggests that while trust is created slowly, it can be destroyed easily, even by a single mistake (Rothbart & Park, 1986). Naturally negative events (failures) decrease trust whereas positive events (successes) increase it. However, as Slovic found, negative events decrease trust far more than positive events serve to increase it. In one survey, he presented his respondents with either the positive or the negative versions of a set of given acts and asked them to rate the impact of each act on their sense of trust (e.g., denying public access versus allowing public access to records of nuclear sites). The results revealed such an asymmetry. Slovic's notion of trust asymmetry seems related to Skowronski and Carlston's idea that negative information is perceived as more diagnostic than positive information and is therefore weighted heavily in the reputation formation process.

In our research the term reputation is used to describe what the decision maker learns online about the advisor through a series of estimation episodes and the ensuing feedback. We suggest that positive and negative reputational information gathered across a series of episodes have a differential impact on the resulting weighting policy. Moreover, in a dynamic situation where the advice quality changes (i.e., declines or improves) the weighting policy may be asymmetrically revised. Thus a poor advisor is discounted more than a good advisor is credited in the process of weighting.

In the following four studies we investigated the discounting hypothesis and our hypothesis regarding the dynamics of reputation formation. We presented

to respondents questions that had real consequences for them as decision makers, since they received a bonus for making accurate judgments. In Studies 1–3 respondents were given advice for free and the principal measure was the weight placed on that advice in their final decision. In Study 4 the principal measure was the respondents' interest in purchasing advisory estimates.

The studies were conducted on a computer due to their interactive nature and shared the following general procedure. In the first phase, respondents were presented with questions and asked to state their estimates. In the second phase, they were presented with the same questions along with estimates made by various advisors (other students). Respondents were then asked to provide their estimates once again. They were free to use the advisors' estimates as they wished. In Study 4 an important procedural variation was introduced in which respondents had to decide whether or not to buy the advisory estimate.

STUDY 1

The first study investigated how people combine their own estimate and the estimate of a randomly drawn judge in response to general-knowledge questions. The experimental procedure was conducted individually on personal computers. Fifteen questions about the dates of historical events (within the last 300 years) were presented sequentially on the computer display screen. As shown in Table 1, in the first phase respondents were shown one question at a time, as shown below, and asked to type in their best estimates for each via the computer keyboard; in addition they were asked to give lower and upper boundaries such that the truth would be included between the limits with a probability of 0.95.

After the first phase was over, the respondents were told that there would be a second phase. In the second phase they were presented with the same set of questions. Now, however, each question was presented along with two estimates: the respondent's own initial estimate and that of an advisor. The respondents then gave a second, possibly revised, estimate for the question. After each question, feedback was given which included the correct answer to that question. This feedback was displayed on the screen immediately after

TABLE 1

Sample Question and General Procedure

Phase 1 (series of 15 questions):
In what year were the Dead Sea scrolls first discovered?
Your best estimate _____ (low estimate _____ high estimate _____)

Phase 2 (same 15 questions repeated):
In what year were the Dead Sea scrolls first discovered?
Your previous best estimate was 1940 (low estimate 1930 high estimate 1960)
Best estimate of advisor K was 1960 (low estimate 1920 high estimate 1975)
Your final best estimate _____ (low estimate _____ high estimate _____)

the final (second) estimate was entered and enabled the respondents to evaluate their own accuracy as well as the accuracy of the advisor's estimates. The respondents were told they would get a bonus depending on their accuracy (see below).

The advisor's estimate was randomly drawn by the computer from a pool of estimates collected in an earlier study in which respondents were merely instructed to provide a best estimate for each question, together with lower and upper boundaries such that the truth would be included between the limits with a probability of 0.95. The advisor varied from one question to the next, with labels such as *J*, *K*, and *P* used to indicate that each estimate came from a different individual. By sampling estimates from pools of data, adequate ecological validity could be maintained. This random mix of estimates had a dispersion and an error corresponding to those that might have been encountered in reality by our respondents when seeking answers to such questions among their peers—undergraduate social science students.

The respondents ($N = 25$) were undergraduate students at the Hebrew University who participated either as part of course requirements or for a flat fee of 12 Israeli shekels. They were all told that they would receive a bonus based on the accuracy of their estimates. In particular, they would receive 1 Israeli shekel ($0.30 at the time of the study) as a bonus for each estimate that had a better than average accuracy score. Altogether they could amass up to 15 shekels in bonus payment. Thus it was in their interest to consider carefully and make the best use of the estimates given to them. The bonus was based on the final estimates (i.e., second phase).

Results

Improving accuracy. Having the advisor's estimate available helped respondents improve their accuracy. The mean absolute error (in years) was lower for the combined estimate (44.2) than for the initial estimate (54.4), $t_{24} = 4.07$, $p < .01$. This gain in accuracy—which is the reason for the broad interest in combining estimates in forecasting—sets the stage for further analysis of the weighting process.

Discounting. How much weight did respondents place on their own estimate? The final estimate can be represented as a weighted combination of the two prior estimates—*own initial* and *advice*—with the weights being proportional to the extent of the shift toward (away from) the advice. Define *weight of own estimate* or $\text{WOE} = |a - f|/|a - i|$ where *a*, *f*, and *i* stand for advice, final, and initial estimates, respectively; WOE is well defined if the final estimate falls between the initial estimate and the advice, as it did in over 95% of the cases. The WOE measure, expressed as a proportion, reflects the weight that a respondent assigns his or her own estimate and also, by the same token, indicates the extent to which the advice is discounted. Thus WOE takes a value of 1.0 if, in making the final estimate, the respondent adheres completely to his or her previous estimate (100% discounting of the advice); WOE is 0 if the respondent shifts completely to the other estimate (0% discounting). Intermediate

values of WOE indicate that positive weights were assigned to both opinions (partial discounting).

Whereas a WOE of 0.50 implies equal weighting, the mean observed WOE 0.71 was significantly higher, $t_{24} = 5.1$, $p < .01$. Respondents placed a higher weight on their own opinion than on the advisor's opinion. This tendency was exhibited by most respondents: 21 of the 25 respondents had a mean WOE greater than 0.5. The interquartile range for respondents' means was from 0.58 to 0.87, and their standard deviation was 0.20.

In another analysis we considered the distribution of all 375 individual trials (25 respondents $\times$ 15 questions). The interquartile range for the individuals trials was from 0.44 to 1.00. The median WOE across 375 individual trials was 0.98. Consider the distribution of WOE. The WOE's were first rounded to the nearest decimal and then grouped into four ranges: 0–.2; .3–.5; .6–.8; .9–1.0. The corresponding percentages falling in each range were as follows: 17, 13, 17, and 53%. These results are instructive and strengthen our conclusion regarding the discounting of advice. One might wonder whether respondents actually knew the correct answers and also knew that they knew them and therefore placed high weights on their own estimates. This notion, however, is not supported by our data, showing that the *exact* correct answer was given on only 4.4% of the trials in phase 1. In contrast, the advice-discounting hypothesis suggests that respondents preferred their own estimates to the advice due to their differential access to information. This hypothesis also allows for the possibility that this preference might be sensitive to the accuracy levels of their own and the advisors' estimates.

Sensitivity to quality. Next we provide a preliminary test of respondents' sensitivity to the accuracy and precision (range size) of their initial estimates and the advice. Yaniv and Foster (1995, 1997) have shown that accuracy and informativeness of interval estimates are two important indicators of the quality of judgments. Accuracy is indicated by absolute error and informativeness by range size. The present study was not designed with correlational analysis in mind and the sample size is too small for proper multiple regression; thus the following results are presented only in an exploratory fashion, to guide further studies. A stepwise regression procedure was performed. The predictors were the mean absolute error of the respondent's own initial estimates, the mean of the respondent's initial ranges, the mean absolute error of the advice set provided to the respondent, and the mean of the ranges in the advice set provided to the respondent. The dependent variable was the mean of the weights of own estimates. The weights were expressed as odds ratios rather than proportions (e.g., WOE of 0.80 was represented as 4:1 and 0.33 as 1:2) and then log-transformed prior to taking the mean. In the stepwise procedure, mean own error emerged as a significant predictor, $R^2 = 0.21$, $F(2, 23) = 6.23$, $p < .05$. No other variable met the 0.15 significance level for entry into the model. Thus decision makers seemed sensitive to the accuracy of their own opinions, presumably due to the accuracy feedback provided online. The mean error of the advice did not emerge as a significant predictor possibly because

its restricted variability—its variance was only one third of that of the mean error of respondents' own estimates. The foregoing regression results are presented here only tentatively, in subsequent studies advice accuracy was controlled systematically.

Discussion

A major conclusion regarding the weighting policies in this study is that decision makers tend to discount advisory estimates egocentrically. Although on average the respondents' accuracy was on a par with the advisors' accuracy (mean absolute errors were 54.4 and 49.9 for respondents and advisors, respectively), respondents placed greater weight on their own judgments. They resolved the discrepancy between their own and the other opinion by adhering to their own opinion and making a token shift to the other opinion. Interestingly, the average weighting of own estimate of 0.71 closely agrees with the results of Harvey and Fischer (1997). In their study, which involved a cue-learning task, respondents shifted their estimates about 20–30% toward the advisor's estimates. Using a time-series forecasting task, Lim and O'Connor (1995) studied how judges integrate a statistical forecast into their own judgmental forecast. Their findings also suggest that respondents discounted the advisory (statistical) forecast.

Whereas discounting is pervasive, it is not the case that respondents discount others' estimates uniformly or indiscriminately. The results of the regression analysis suggest that respondents' weighting policies were sensitive to the quality of their own estimates. The following three studies were designed to test the boundary conditions on the discounting of advice and trace the reputation formation process. In Study 2, we systematically varied the quality of the advice (good versus poor) and tested whether the respondents' weighting policy is appropriately sensitive to the differences in quality. In Studies 3 and 4 we assessed how rapidly respondents revise the advisors' reputation in response to changes in the quality of the advice.

STUDY 2

In this study, we investigated the weighting policies that are formed in response to either good or poor advice. Each respondent in the present study was paired with either a good or a poor advisor. The good and poor advisors were two actual respondents who participated in previous studies and who were selected on the basis of their performance. Their estimates were given as advice in this study.

In addition to testing the effect of advice quality, we also tested the effect of getting online feedback on the formation of a weighting policy. In particular, we assessed how well respondents distinguish between good and poor advice in the presence and absence of accuracy feedback. The online feedback consisted of the correct answer and was given immediately after each final estimate was entered, just as in Study 1. Clearly, the provision of such feedback can facilitate

learning about the quality of the advice, and hence the formation of reputation. In many realistic situations, however, feedback is not available online, and so it is important to find how critical the provision of feedback is to reputation formation.

Method

The procedure included two phases, as in Study 1. In the first phase the respondents ($N = 80$) were asked to produce estimates. In the second phase they received the same list of questions along with advice and were instructed to provide their final estimates. The advice in this study consisted of the estimates made by one of two preselected advisors. Half of the respondents were assigned to the good advisor condition. The advice they received consisted of the estimates made by the most accurate individual in a previous study, selected from a sample of 50 (the mean absolute error for that respondent was 24 and the hit rate, namely the proportion of questions for which her range estimate included the true answer, was 93%). The other half of the respondents were assigned to the poor advisor condition, where the advice consisted of the estimates made by the least accurate individual in the same sample (mean absolute error 103 and hit rate 13%).

Quality of advice (good, poor) was crossed with the feedback factor (with, without). Half the respondents in each advisor condition received online outcome feedback, whereas the other half did not receive such feedback. In the feedback condition the correct answer to each question was presented on the screen after the respondent had entered the second (final) estimate. This feedback enabled them to assess the magnitudes of errors (advisory and own) and check on whether either range included the correct answer (i.e., hit). The two feedback types were in high agreement with respect to each advisor—the good advisor had a high hit rate as well as a low mean absolute error and the reverse was true of the poor advisor.

In either condition respondents were not told anything about the quality of the advice. They were simply told (as they were in Study 1) that they would receive a bonus of 1 Israeli shekel ($0.30 at the time of the study) as a bonus for each estimate that had a better than average accuracy score. They could earn up to 15 shekels in bonus payment. Thus it was in their interest to consider carefully and make the best use of the estimates given to them.

Results

Table 2 suggests that WOE differed as a function of the quality of the advisor. Analysis of variance on WOE with *advisor quality* (good, poor) and *feedback* (with, without) as factors revealed an effect of advisor, $F(1, 76) = 59.67$, $p < .001$, no effect of feedback, but an interaction effect, $F(1, 76) = 5.14$, $p < .05$. Thus, respondents' WOE was significantly higher when paired with a poor advisor than with a good one. The interaction shows that the effect of advisor quality was enhanced when feedback was given—with feedback, respondents

TABLE 2

Mean Weight of Own Estimate (WOE) in Study 2 as a Function of Advisor Quality and Feedback

	Feedback	
Advisor quality	With	Without
Poor advisor	0.76	0.73
Good advisor	0.42	0.54

Note. The pooled standard deviation (square root of MSE) in this table is 0.26.

could better differentiate the quality of the advice. There was significant simple effect of advice quality, both in the feedback condition, $F(1, 76) = 16.60$, $p < .05$, and in the no-feedback condition, $F(1, 76) = 5.18$, $p < .05$.

Recall that respondents were not told anything directly about the quality of the advice they got. They could infer the quality from the characteristics of the estimates and the feedback (when it was given). We performed two regression analyses (one for each feedback condition) to assess the relationship between the weighting policies and the error and range of the estimates. We note that the following analyses were only cursory and meant merely to supplement the analysis of variance above. As in Study 1, two stepwise regressions were done on the mean (log transformed) weight variable (expressed as odds ratio rather than proportion) with the four predictors defined in Study 1 (i.e., mean absolute error of the respondent's own initial estimates, mean of the respondent's initial ranges, mean absolute error of the advisor's estimates, and mean of the ranges of the advisor's estimates). In the no-feedback condition the respondent's own mean error and the advisor's mean range were significant predictors, $R^2 = 0.38$, the F-values for these predictors were 11.3 and 9.1, $p < .01$, $df = 37$ for the error term. No other variable met the 0.15 significance level for entry into the model. In the feedback condition, the respondent's own mean error and the advisor's mean range were significant, $R^2 = 0.74$, the F-values were 75.8 and 10.9 for the two predictors, $p < .001$, $df = 37$ for the error term. Other variables did not meet the 0.15 significance level for entry into the model. These results perhaps explain respondents' sensitivity to the quality of the advice even in the no-feedback condition. They are consistent with the idea that the advisor's ranges were considered in weighting the advice. This explanation and others are further discussed in the following section.

Discussion

The respondents' weighting policies were sensitive to the quality of the advice. Respondents formed valid impressions of the advisors although they were not told anything directly about their quality. In the feedback condition, decision makers could rely on the outcome feedback, which was directly associated with the stimulus question (i.e., contiguous in time and space). It is intriguing,

however, that even in the absence of feedback, decision makers were sensitive to the quality of the advice, as their weighting policies show. There are two explanations for this finding. First, respondents relied on the ranges of the estimates as a cue to accuracy. Indeed, the width of an interval estimate is a partially valid predictor of accuracy (Yaniv & Foster, 1997). In the present study, range was used as cue in weighting the advice.

Our second explanation is that respondents could perform plausibility checks on the advice to evaluate its quality, even in the absence of feedback. They might have recognized that a particular poor estimate was out of bounds, even if they themselves could not generate a correct one. For instance, a person who cannot produce the exact date of the U.S. Declaration of Independence might still be able to infer that the provider of an estimate such as "1450" is poorly knowledgeable. In a different vein, upon receiving a good piece of advice, respondents might have been reminded of information that they previously failed to recover themselves from memory, although it was stored there (Yaniv, Meyer, & Davidson, 1995). Respondents might build reputation for an advisor in the absence of feedback if they can perform such plausibility checks even on only a few test cases.

Consider now the effect of advice quality on the weighting policies. The weighting policies in the good- and poor-advisor conditions differed markedly from one another (WOE averaged 0.48 and 0.74, respectively). Consider equal weighting (50:50) as a baseline relative to which weighting policies can be assessed. Whereas poor advice was clearly discounted, one might argue that even the good advisor was discounted to some extent. Respondents weighted the good advice and their own opinion about equally, although the good advisor (mean absolute error = 24.0) was better than 98% of the respondents in the present study (mean absolute error = 63.5). Clearly they might have assigned greater weight to the good advisor's estimates.

It is interesting that the mean WOE in the poor advice condition (0.74) was close to the mean WOE in the first study (0.71), where the advice was randomly drawn and hovered about average quality. Is it the case that respondents either neglect or are reluctant to distinguish between an about-average advisor and a poor advisor? We interpret this finding to be in line with our hypothesis on reputation formation. The about-average advisor of Study 1 offered both good and poor advice. The notion of trust asymmetry predicts that cases of poor advice would loom large in the minds of decision makers. In other words, an average advisor is deemed poor due to asymmetry in reputation formation. The third study was designed specifically to test this implication of trust asymmetry by carefully monitoring how the weighting policy is updated as the quality of the advice changes.

STUDY 3

Studies 1 and 2 show that, in establishing a weighting policy, respondents are prone to discount advice. We have suggested that a reputation for being a

good advisor is more easily lost than gained (Slovic, 1993). Reputation building may be likened to biking uphill; that is, it is hard to start forward (gain) but easy to roll backward (lose). To test this idea, Study 3 evaluated how decision makers' weighting policy is revised as the quality of advice changes along the way. In preparing this study, pools of answers from previous studies were used. Each pool of answers to a given question was ranked in terms of absolute errors. The good advice consisted of estimates that ranked in the top 10% of the pool (mean absolute error = 4.7, while the hit rate, namely the proportion of questions in which the range estimate included the true answer, was 90%); the poor advice consisted of estimates that ranked at the bottom 10% of the pool of answers to that question (mean absolute error 63.0, hit rate 5%). The average-quality advice presented in the second part were estimates whose errors were about the median (mean absolute error 33.2, hit rate 38%).

We used the same two-phase procedure as in Studies 1–2, introducing small and large changes. The minor change is that 21 questions were now included. The major change is that the quality of the advice in phase 2 changed along the way in the following manner. The head of the sequence (first k trials) included either good or poor advice. The tail of the sequence (remaining trials) included only average-quality advice. In the first-poor-then-average condition, the quality of the advice was poor at the head of the sequence (first k trials) and then switched to average at the tail of the sequence (remaining trials). In the first-good-then-average condition, the quality of advice was good at the head of the sequence and then switched to average in the tail trials. The head of the sequence consisted of either 3 or 9 trials. In sum, two factors were crossed in this study: the *advice quality* at the head of the sequence (either good or poor in the first k trials) and the *length* of that first part (k = 3 or 9 trials).

Our goal was to evaluate how respondents react to changes in the quality of the advice. An interesting question is whether the respondents in the two groups would converge onto the same or different weighting policies on the tail trials, depending on their early experience (i.e., the advice quality at the head of the sequence). Second, we wondered whether the average advice would be weighted closer to the good or the poor advice. If the formation of reputation is such that it is gained with difficulty but easily lost, then we would expect the weighting of average advice to be a large step down from the weighting of good advice. In contrast, the weighting of average advice should be only a small step up from the weighting of poor advice.

Feedback was given on all trials—respondents received the correct answer to each question online after they had entered their final estimate. As in Studies 1 and 2, respondents were *not* told anything about the quality of the estimates they were given. They were simply told that they should make the best use of the advice because they could earn 1 Israeli shekel ($0.30 at the time of the study) per each final estimate that had a less-than-average error. Altogether they could earn 21 shekels in bonus payments. There were 104 respondents in this study, who were divided equally among the four conditions.

Results

Figure 1 shows respondents' WOE's separately for the head of the sequence (block 0) and the tail of the sequence (blocks 1–3). Respondents were highly sensitive to the quality of the advice: the mean WOE was 0.82 when paired with a poor advisor and 0.41 when paired with a good advisor. The first analysis of variance was performed only on the data from the head of the sequence (initial k trials), with WOE as the dependent variable and advisor quality (good, poor) and length (9 = long, 3 = short) as factors. There was a significant effect of quality, $F(1, 100) = 86.0$, $p < .001$, and no significant effects of either length or interaction.

The results (Fig. 1) for the tail of the sequence (subsequent 12 trials) were broken down into three blocks of four trials each (blocks 1–3). In the $k = 3$ condition we plotted trials 4–15, whereas in the $k = 9$ condition we plotted trials 10–21. A $2 \times 2 \times 3$ repeated measures analysis of variance was performed on WOE with quality (good, poor), length (3, 9), and block number (1–3) as factors. There were no significant effects of either quality or length on WOE across blocks 1–3, $F(1, 100) < 2$.

To detect block by block changes, we also report the analysis results separately for each block as a function of the head of the sequence. In block 1 we found a significant effect of advice quality, $F(1, 100) = 5.34$, $p < .05$, but no effect of length and no interaction effect, the Fs were 0.3 and 1.3, respectively. In block 2 there were no effects of advice quality, length, or interaction, $F < 2.1$. Similarly, in block 3 there were no effects, $F < 1.9$.

Discussion

Recall that the quality of advice was either good or poor during the head trials (block 0), and average in the tail trials (blocks 1–3). The advice quality had a strong effect on WOE in block 0 (head trials). There was also a significant effect on WOE in block 1—a lingering reputational influence of block 0. However, this lingering influence was virtually erased in blocks 2 and 3. Respondents updated their weighting policy rapidly, within a few trials (each block

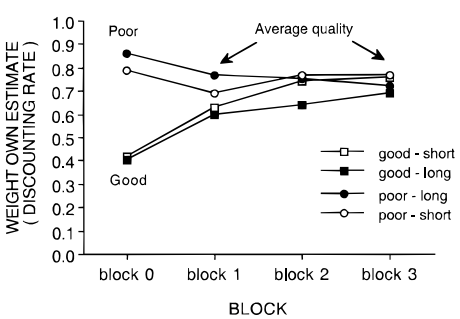

FIG. 1. Mean WOE (Study 3) is shown for the initial K trials (block 0) and the subsequent 12 trials (blocks 1–3) as a function of advice quality and length of block 0.

included four trials). The mean WOE was relatively stable in blocks 2 and 3, hovering in the vicinity of 0.72 in all four conditions, as can be seen in Fig. 1. The respondents in both conditions converged to roughly the same weighting policy. For instance, respondents who initially received good advice reverted to a mean WOE of 0.71 in blocks 2 and 3—a relatively large step up in discounting (0.41 to 0.71). Respondents who initially received poor advice also reverted to a mean WOE of 0.74—a relatively small step down in discounting (0.82 to 0.74). To summarize, the weighting policy is rapidly updated following changes in quality. This pattern of updating is consistent with the notion that reputation is more easily lost than gained.

STUDY 4

A central finding from the previous studies is that reputation is formed rapidly in the respondents' minds and also updated quickly when the advice quality changes. The evidence was based on the weighting policies used by decision makers, specifically, on the WOE measure which indicated the extent to which the advice was discounted. In realistic situations, reputation certainly affects the weighting of advice. However, reputation may also affect whether the advice would be solicited in the first place. The goal of the present study was to test the effect of advisor reputation on the willingness to buy advice.

The general procedure was as follows. During a period of several preliminary trials, advice was given for free so reputation could be formed. Then, on subsequent trials, respondents were told they could purchase advice at a fixed cost. Thus, on each of those trials they were to decide whether or not to purchase an estimate for a fee that would be deducted from their final payment. The dependent variable was their buying rate of advice and was measured as a function of the initial reputation formation period. This procedure simulates realistic situations in which decision makers actively decide whether or not to seek advice. A decision maker's attitude toward an advisor is revealed in their willingness to invest the resources (efforts, time, money, etc.) that are required to solicit advice from that person.

We measured the buying rate of advice as a function of its perceived quality. We therefore fixed the cost of advice and the bonus for giving a good estimate. A preliminary study indicated that a fee in the neighborhood of a quarter of a shekel (vis-à-vis a potential bonus of 1 shekel) makes the purchase of advice moderately attractive (advice was purchased roughly from one quarter to one half of the time in this pilot study). Since we were only interested in the course of reputation formation, we did not manipulate the cost of the advice or the monetary incentives for an accurate estimate (cf. Connolly & Thorn, 1987).

Method

This study, like the previous ones, was conducted on personal computers and included two phases. In the first phase, respondents were presented with the same 21 questions as in Study 3 (in a randomized order). The second phase

consisted of two parts: the head of the sequence, in which reputation could be formed, and the tail of the sequence, in which buying rate was measured. For each of the first *k* questions, advice was automatically presented for free. (The value of *k* was either 3 or 9.) For the remaining trials (the tail of the sequence), the opportunity to purchase advice was given as an option. The cost of advice was fixed at 0.25 shekel per estimate, and a bonus of 1 shekel was offered for each final estimate that had a lower than average error. On each trial, respondents were shown their own initial estimate from phase 1 and asked whether they wanted to buy an estimate.

If they answered "yes" ("y" key on the computer keyboard) an advisor's estimate was shown to them along with the request to enter a final estimate. If they answered "no" ("n" key), then they were told to enter their final estimate. Respondents were given feedback at the end of each trial, immediately after entering the final estimate. They were told that the fee per piece of advice would be deducted from their final pay, whether or not they earned the bonus. Because they were paid a flat fee for participation (see below), the total payment at the end of the experiment was always positive.

Two experimental factors were varied in the study. The first was the quality of the advice (good or poor). The selection of good and poor advice was based on the same criteria as in Study 3. In one condition good advice was given on all 21 trials. In the second condition poor advice was given on all 21 trials (the mean accuracy scores for these advisors were the same as in Study 3). The second factor was the length (*k*) of the free experience period, either short (3 trials) or long (9 trials). Thus, four conditions were created by crossing these two factors.

Ninety-two respondents were randomly assigned to the four conditions (23 in each condition). As in the previous studies, they were *not* told anything about the quality of the advice. They were merely told that they should make the best use of the advice because they could earn 1 shekel per each final estimate that had a less-than-average error. Thus in addition to a flat fee of 12 shekels, they could also earn up to 21 shekels in bonus fees minus the fees paid for advice.

Results and Discussion

Figure 2 shows the buying rate (broken into three blocks of trials) as a function of advice quality (good, poor) and the length of the preliminary period (long, short). Analysis of variance was done on the advice-buying rate with two between-subjects factors (advisor quality and length of preliminary experience period) and one within-subjects factor (block number). We found a significant effect of advice quality, $F(1, 88) = 101.8$, $p < .001$, no significant effect of the length of the preliminary period, $F(1, 88) = 2.85$, $p > .05$, and no block effect, $F(2, 176) = 2.67$, $p > .05$. None of the interactions was significant.

Respondents clearly established fundamentally different impressions of the good and the poor advisors on the basis of the preliminary trials. They purchased advice far more often from the good advisor (about 70% of the time)

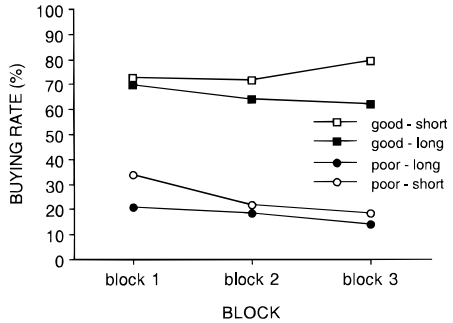

FIG. 2. Mean buying rates (Study 4) in three blocks (6 trials in each) as a function of the quality and length of the sequence of preliminary trials.

than from the poor advisor (about 20%). In comparison with this large effect of quality, the length of the initial learning period (3 vs 9 trials) had no statistical effect on the buying rate. This implies that reputation was formed rapidly, presumably within three trials. Finally, the buying rates remained relatively stable across blocks 1–3, with occasional fluctuations that appear to be random.

The buying rate was analyzed in this study, but WOE was not. The WOE measure could be calculated only when respondents chose to buy advice, thus subjecting it to severe selection problems, since poor advice was bought far less frequently than good advice.

GENERAL DISCUSSION

A fundamental question in the study of behavioral decision making is: What does the decision maker do with the available information? In the present research we considered the use of advisory estimates. We have argued that taking advice is a central component of realistic decision making since decision problems often do not come as complete, self-contained textbook problems. Instead, decision makers engage in interactive social and cognitive processes to build the informational relevant basis. In particular, they solicit opinions from worthy advisors, assess their merit, and then combine them. Our framework for understanding decision makers' use of advice involves two theoretical concepts that motivated the studies and serve to explain the findings. The first is egocentric discounting of another's opinion and the second is reputation formation for advisors. Briefly, advice is discounted relative to one's own opinion, and advisors' reputation is rapidly formed and asymmetrically revised. We review the evidence for these concepts, trace their theoretical origins, and point out some of their implications.

Discounting Advice

In three experiments we considered decision makers' weighting policies. In the first study both the decision makers and the advisors were undergraduate social science students and their estimates were almost equally good in terms of accuracy. In the analysis of the weighting policies we found that decision

makers egocentrically discounted the advisor's opinions relative to their own. The average weight on own estimate (WOE = 0.71) was significantly greater than 0.50. In Study 2 the WOE was 0.74 when the advice was poor and 0.48 when the advice was produced by a good advisor. It is notable that decision makers assigned almost the same weight to the good advisor as to themselves, even though the good advisor was superior to nearly all of them (in terms of lower absolute error). One might argue therefore that even the good advisor was discounted.

Converging evidence for the discounting of advice comes from the laboratory as well as field studies. In studies by Harvey and Fischer (1997) respondents were first trained in a cue-learning task. In the test session, the respondents made initial estimates and then final estimates on the basis of a recommendation from an advisor. The advisors had supposedly received different amounts of training in this artificial environment. Harvey and Fischer found a shift in judgment of about 20–30% toward the advice, which corresponds to a WOE of about 70–80%—a result consistent with ours. Using a time-series forecasting task, Lim & O'Connor (1995) studied how judges integrate a statistical forecast into their initial judgment-based forecast. Explicit outcome feedback was provided on each trial. Their findings show that respondents assigned about double the weight to their initial forecast than to the statistical forecast.

The phenomenon that individuals stick to their initial opinions has been documented in at least one professional setting. In his literature review on the impact of genetic counseling, Kessler (1989) concludes that genetic counseling does not produce dramatic changes in counselees' reproductive decisions. The best predictor of the postcounseling reproductive decision is the counselee's precounseling intention (Sorenson, Scotch, Swazey, Wertz, & Heeren, 1987).

It may be fruitful to tie our hypothesis on the discounting of advice to the literature on the public's perception of risks (such as environmental, health-related, and technologically-induced risks). A recurring finding is that the experts and the public differ in their perception of such risks, thus creating difficulties in the implementation of public policy (Flynn, Slovic, & Mertz, 1993; Slovic, 1987, 1993). Experts' risk communication could be construed in our framework as advising individuals' daily decisions regarding the safety measures they need to take against various types of risks (e.g., using seatbelts in a car's back seat) and how much to invest in related efforts (e.g., whether to join a campaign for the removal of a safety hazard). The alleged skepticism toward expertise in this area can be viewed as partial discounting of expert advice. This sort of disagreement could arise even if lay people consider experts' opinions, but also assign nonzero weight to their own opinion.

We suggest that advice-discounting results from decision makers' differential access to the underlying justifications for each opinion. Note first that a stated opinion is an external expression that summarizes one's own relevant internal knowledge base. The decision maker is privy to the reasons supporting his or her own estimate as well as to the strength of those reasons, but is not privy to the internal network of reasons underlying the advisor's summary opinion.

A common assumption of the cognitive approach is that the weight placed on a certain judgment (hypothesis) depends on the evidence that could be recruited to support that judgment (Tversky & Koehler, 1994). Therefore, in advice-taking situations, the decision maker's view and the advisor's opinion are not on a par. Other things being equal, decision makers discount the advisor's opinion relative to their own due to their differential information about the respective underlying justifications.

The notion that one has privileged access to one's own reasoning but not to others' reasoning provides a parsimonious account of the differential weighting of one's own and others' opinions. Moreover, it paves the way for the explanation of the effects of quality on the weighting policies (via reputation formation). Other, less parsimonious explanations of the egocentric discounting could be based on the notions of overconfidence, perseverance bias, and confirmation bias. There is evidence demonstrating that individuals tend to be overly confident about their own predictions and estimates (Tversky & Kahneman, 1974; Lichtenstein, Fischhoff, & Phillips, 1982). Individuals also tend to retrieve from memory evidence consistent with their estimates rather than evidence refuting their estimates (Hoch, 1985; Strack & Mussweiler, 1997). In addition, they tend to maintain internal consistency by discounting the validity of evidence contradicting their opinion rather than by revising their opinion (Lord, Ross, & Lepper, 1979). Explanations in terms of these factors, however, would require more assumptions and thus would be less parsimonious.[1]

Nevertheless, we do not exclude the possibility that additional factors could at times exacerbate the egocentric weighting. For example, if a decision maker's opinion is made public or if a change of opinion is perceived as ego-threatening, impinging on self-perception (e.g., political or religious beliefs), then additional forces may come into play, further heightening the egocentric tendency to discount advice. However, the conditions that typically create and enhance commitment to one's own position were not present in our procedure, which involved neutral questions, was conducted in isolation in front of a computer terminal, and rewarded for accuracy. Moreover, an explanation in terms of the need to protect consistency does not readily explain respondents' sensitivity to the changing quality of advice. In contrast, the hypothesis that respondents rely on assessments of their internal evidence and the external feedback readily explains their sensitivity to advice quality. We therefore believe that the most parsimonious explanation of the advice discounting phenomenon is based on the notion that the decision maker has privileged access to his or her own reasons but only limited (if any) access to the reasons supporting the advisor's opinion.

[1] Decision makers' shift toward some advice might also depend on whether they had an opportunity to express their own, independent opinion prior to getting advice. In our studies individuals always stated their estimate prior to getting advice. One could test the hypothesis that individuals who express their opinion only *after* being shown a piece of advice would shift relatively more toward the advice.

Reputation Formation

Whereas advice discounting is pervasive, we do not claim it is uniform in magnitude or applied indiscriminately across all situations and types of advisors. In fact, respondents in our studies showed high sensitivity to the relative quality of the advice compared with their own knowledge, as the regression analyses indicate. We suggest that decision makers' weighting policies are affected by advisors' reputation and that the formation and revision of reputation is subject to an important asymmetry.

In particular, good reputation is more easily lost than gained. The mean WOE in the poor advice condition of Study 2 (0.74) was close to the mean WOE in Study 1 (0.71), where the advice was randomly drawn and hovered around average quality. In Study 3 the quality of the advisor's estimates either declined from good to average or improved from poor to average. When the quality of advice declined, respondents sharply boosted their discounting of advice (WOE increased from 0.41 to 0.72). In contrast, when the quality of the advisor's estimates improved, respondents reduced their discounting level only slightly (WOE decreased from 0.81 to 0.74). These results are consistent with Slovic's notion of trust asymmetry (1993) and the negativity effect in impression formation mentioned in the introduction. The results also suggest that good advisors are in a class of their own, whereas average and poor advisors are not well differentiated. What is the basis for these findings?

We have already mentioned Skowronski and Carlston's (1989) recent explanation for the negativity effect which is based on the notion that negative information about an actor is viewed by the social perceiver as more diagnostic than positive information. Related to that is Rothbart and Park's (1986) suggestion that more confirming evidence is required to establish a favorable trait (such as trustworthiness) than to establish an unfavorable trait in a person. Slovic (1993) also mentions that negative information receives greater weight in judgment because it is cognitively more salient and emotionally more loaded.

We might add here the conjecture that risk aversion also contributes to the asymmetry in reputation formation and sheds light on why average and poor advisors are treated almost alike whereas good advisors are differentiated. Risk aversion here implies that the downside potential of adopting poor advice looms large (relative to the benefit of good advice). An average advisor is someone whose advice, by definition, hovers about the average, being sometimes good and sometimes poor. An average advisor might then be viewed as more damaging than helpful. This might explain why the average advisor in our results seems closer to the poor advisor than to the good advisor (in terms of WOE). The risk aversion explanation of the reputation formation asymmetry is offered here only tentatively, but it seems worthy of further testing.

The speed with which reputation is formed emerges in our studies as an important phenomenon in its own right. In Study 4 decision makers had to decide whether or not to purchase estimates from an advisor whose advice

they had received on previous trials. A few (about three) preliminary trials sufficed to guide these buying decisions powerfully. Similarly, in Study 3, when the quality of the advice either dropped or improved along the way, decision makers also updated their weighting policy fairly quickly, within a few trials after the shift.

Reputation formation is closely related to both impression formation and learning processes. On the one hand, a reputation is founded on impressions. In this respect, our studies bear similarity to impression formation studies, which typically involve judgments about hypothetical personality profiles and tend to emphasize the role of heuristic processes, such as confirmation bias, primacy effect, and negativity effect (Fiske, 1980; Hamilton & Huffman, 1971; Skowronski & Carlston, 1989). On the other hand, our studies involved learning since our respondents made judgments about real events, received outcome feedback, and earned a bonus depending on their knowledge and judicious use of the advice. Hence fine-tuning of the weighting policies is to be expected.

The sequences of 15 to 24 trials in our studies may seem short compared to the dozens (or even hundreds) of trials administered in some learning experiments. However, informal observation suggests that an advisor's reputation might be formed on the basis of a small number of encounters. Decision makers (consider, for instance, a journal editor, a business person, or a patient) may not pursue a relationship with an advisor (e.g., a referee to a journal, a business consultant, or a physician) who has failed on several previous occasions; such an advisor may have little or no opportunity to improve his or her reputation. Short sequences of trials may be appropriate for studying the dynamics of advice taking in social and professional situations that are characterized by a brief window of opportunity during which an advisor can rapidly gain or lose a good reputation. Reputation formation seems to be a heuristic process that is potentially reactive toward negative information, subject to the effect of recent trials, and based on quick generalizations made on the basis of few data points. Future work could investigate how reputation formation is affected when learning occurs across longer sequences of trials.

Finally, we note in passing that reputation might be formed even in the absence of feedback as we have seen in Study 2. For one thing, decision makers could use cues to accuracy, such as the uncertainty of the estimate (Yaniv & Foster, 1997). Moreover, decision makers could sometimes pass judgment on the plausibility of a piece of advice, even when they themselves were uncertain about the truth or correct outcome. A piece of poor advice could, on occasion, be recognized as such since one does not always need to know the exact answer to a question in order to judge that someone else's answer is patently wrong or even ludicrous. Similarly, a piece of good advice could act as a retrieval cue, reminding respondents of information that they had previously failed to recover themselves from memory (Yaniv, Meyer, & Davidson, 1995). Valid plausibility checks might not be possible on each trial. Our point, however, is that even a few test cases could lead respondents to establish a reputation for the advisor.

Concluding Remarks

The research on giving and taking advice highlights the role of social processes in decision making (Sniezek & Buckley, 1995; Yates, Price, Lee, & Ramirez, 1996). Physicians, weather forecasters, genetic consultants, and lawyers, to name only few of the professionals who give advice, are all in the business of communicating their uncertain estimates to other individuals. We therefore need to understand better how advisory estimates are used by decision makers. In the present work we focused on how advice is weighted as a function of the advisor's reputation for accuracy. But we envision additional bases for discounting and reputation formation. For instance, discounting might be triggered by lack of informativeness of the advice (Yaniv & Foster, 1995, 1997) and poor reputation may arise from an advisor's failing to act in good faith. The role of advice taking in decision making thus appears to be a broad topic and a challenge for further research.

REFERENCES

Ashton, A. H., & Ashton, R. H. (1985). Aggregating subjective forecasts: Some empirical results. *Management Science*, **31,** 1499–1508.

Connolly, T., & Thorn, B. K. (1987). Predecisional information acquisition: Effects of task variables on suboptimal search strategies. *Organizational Behavior and Human Performance*, **39,** 397–416.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, **34,** 571–582.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, **13,** 171–192.

Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin*, **84,** 158–172.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, **38,** 889–906.

Flynn, J., Slovic, P., & Mertz, C. K. (1993). Decidedly different: Expert and public views of risks from a radioactive waste repository. *Risk Analysis*, **13,** 643–648.

Graham, J. R. (1996). Is a group of economists better than one? Than none? *Journal of Business*, **69,** 193–232.

Hamilton, D. L., & Huffman, L. J. (1971). Generality of impression-formation processes for evaluative and nonevaluative judgments. *Journal of Personality and Social Psychology*, **20,** 200–207.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment and sharing responsibility. *Organizational Behavior and Human Decision Processes*, **70,** 117–133.

Hill, G. W. (1982). Group versus individual performance: Are $N + 1$ heads better than one? *Psychological Bulletin*, **91,** 517–539.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11,** 719–731.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, **21,** 40–46.

Kessler, S. (1989). Psychological aspects of genetic counseling: A critical review of the literature dealing with education and reproduction. *American Journal of Medical Genetics*, **34,** 340–353.

Libby, R., & Blashfield, R. K. (1978). Performance of a composite as a function of the number of judges. *Organizational Behavior and Human Performance*, **21,** 121–129.

Lichtenstein, S., Fischhoff, B., & Phillips, P. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Lim, J. S., & O'Connor, M. (1995). Judgmental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, **8,** 149–168.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, **37,** 2098–2109.

Rothbart, M., & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology*, **50,** 131–142.

Seife, C. (1996). Web pits one against the world. *Science*, **273,** 1166.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, **105,** 131–142.

Slovic, P. (1987). Perception of risk. *Science*, **236,** 280–285.

Slovic, P. (1993). Perceived risk, trust, and democracy: A systems perspective. *Risk Analysis*, **13,** 675–682.

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in Judge-Advisor decision making. *Organizational Behavior and Human Decision Processes*, **62,** 159–174.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, **43,** 1–28.

Sorenson, J. R., Scotch, N. A., Swazey, J. P., Wertz, D. C., & Heeren, T. C. (1987). Reproductive plans of genetic counseling clients not eligible for prenatal diagnosis. *American Journal of Medical Genetics*, **28,** 345–352.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, **73,** 437–446.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185,** 1124–1131.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101,** 547–567.

Wallsten, T. S., Budescu, D. V., Erev, I., & Diederich, A. (1997). Evaluating and combining subjective probability estimates. *Journal of Behavioral Decision Making*, **10,** 243–268.

Winkler, R. L., & Poses, R. M. (1993). Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Management Science*, **39,** 1526–1543.

Yaniv, I. (1997). Weighting and trimming: Heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes*, **69,** 237–249.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness tradeoff. *Journal of Experimental Psychology: General*, **124,** 424–432.

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision making*, **10,** 21–32.

Yaniv, I., & Hogarth, R. M. (1993). Judgmental versus statistical prediction: Information asymmetry and combination rules. *Psychological Science*, **4,** 58–62.

Yaniv, I., Meyer, D. E., & Davidson, N. S. (1995). Dynamic memory processes in retrieving answers to questions: Recall failures, judgments of knowing, and acquisition of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21,** 1509–1521.

Yates, J. F., Price, P. C., Lee, J., & Ramirez, J. (1996). Good probabilistic forecasters: The "consumer's" perspective. *International Journal of Forecasting*, **12,** 41–56.

Zajonc, R. B. (1962). A note on group judgments and group size. *Human Relations*, **15,** 177–180.

Zarnowitz, V. (1984). The accuracy of individual and group forecasts from business and outlook surveys. *Journal of Forecasting*, **3,** 11–26.

Received July 27, 1998; published online August 8, 2000