Using advice from multiple sources to revise and improve judgment

Article *in* Organizational Behavior and Human Decision Processes · February 2007

DOI: 10.1016/j.obhdp.2006.05.006 · Source: RePEc

Ilan Yaniv
Hebrew University of Jerusalem

Maxim Milyavsky
Ono Academic College

Available online at www.sciencedirect.com

ScienceDirect

Organizational Behavior and Human Decision Processes 103 (2007) 104–120

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES

www.elsevier.com/locate/obhdp

Using advice from multiple sources to revise and improve judgments ☆

Ilan Yaniv *, Maxim Milyavsky

Department of Psychology, Hebrew University, Mt. Scopus, Jerusalem 91905, Israel

Received 21 July 2005
Available online 28 July 2006

Abstract

How might people revise their opinions on the basis of multiple pieces of advice? What sort of gains could be obtained from rules for using advice? In the present studies judges first provided their initial estimates for a series of questions; next they were presented with several (2, 4, or 8) opinions from an ecological pool of advisory estimates (Experiment 1), or with artificial advice (Experiment 2); finally they provided their revised estimates. Descriptive analyses of their revision process revealed that they egocentrically trimmed the opinion sets such that opinions distant from their own were greatly discounted. Normative analyses suggest that they gained substantially from the use of advice, though not optimally, due to their self-centered utilization of the advice. The results are discussed in connection with theories of belief revision and attitude change, with an emphasis on decision-makers' strategies for coping with conflicting opinions and the appropriateness of discounting distant or dissenting opinions. Prescriptive implications for the utilization of advice are also considered.

Keywords: Judgment under uncertainty; Combining opinions; Utilizing advice; Decision making; Egocentric judgment

Introduction

It is common practice to solicit other people's opinions prior to making a decision. An editor solicits two or three qualified reviewers for their opinions on a manuscript; a patient seeks a second opinion regarding a medical condition; a consumer considers the "word of mouth" of a dozen people for guidance in the purchase of an expensive product. All these situations involve decision-makers in the task of combining multiple opinions to revise their own (Sniezek & Buckley, 1995). The rationale for soliciting advice is straightforward. Real-life decisions are often not self contained—the range of possible options for choice and their descriptions are often not fully specified. Decision-makers solicit advice to gain information, help them frame their decisions, refine their preferences, and create options beyond those available to them at the moment. At times, people may seek advice for other reasons, such as self-affirmation or for sharing responsibility due to concerns about accountability to others (Kennedy, Kleinmuntz, & Peecher, 1997). Such social reasons are also rooted in the belief that getting advice should ultimately be beneficial to the decision process.

We explore the following paradigmatic situation here. A decision-maker first forms an initial opinion about some issue. Then she receives multiple advice (e.g., two to eight opinions generated by other judges) on the basis of which she revises her initial opinion. We investigate two fundamental issues—first, the influence of advice on decision-makers' final opinions and the *revision rules* they employ in combining the opinions; and second, the *benefits* of using advice, specifically, both the potential

☆ This research was supported by Grants Nos. 822/00 and 344/05 from the Israel Science Foundation to the first author.

* Corresponding author.
E-mail address: ilan.yaniv@huji.ac.il (I. Yaniv).

0749-5978/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2006.05.006

and the actual gains that could be obtained by using advisory opinions.

In these experiments, we consider perhaps the simplest form of advice use, namely getting pieces of information (numerical estimates) from outside parties and using them to update one's own view. As simple as it is, numerical advice has an important function in decisions. Experts such as physicians, weather forecasters, and business consultants often communicate their forecasts and uncertain estimates to others facing decisions. In addition, the use of numerical estimates has certain methodological advantages, primarily the ability to quantify straightforwardly the influence and benefits of using the advice (i.e., participants' revision policies and accuracy gains).

A key issue in integrating advice from multiple sources involves the difficulty of dealing with conflicting advice (Sniezek & Buckley, 1995). Dissenting opinions pose a challenge to the decision-maker, as when two advisors recommend one course of action, while a third one recommends another (Harries, Yaniv, & Harvey, 2004). On top of advisor disagreement, decision-makers need to reconcile potential disagreements between the advisors' and their own opinions (i.e., self vs others).

The present research seeks answers to the following questions: How do people resolve potential conflicts between their own opinions and a sample of advisors' opinions? How do they weigh a dissenter's opinion vis-à-vis the "consensus opinion"? What might be a good strategy for combining multiple pieces of advice? The main contribution of this research is in bringing together three issues. First, it involves multiple (rather than single) advice. Second, it is focused on the process of revising one's *prior* opinion based on advice (rather than purely combining estimates). Third, we conducted a parallel investigation of descriptive and normative issues, where the normative results provide useful benchmarks for assessing decision-makers' performance. This allows us to assess how good or adaptive people's revision rules are.

Our findings lead us to several conclusions. We find that decision-makers gain substantially from the use of multiple pieces of advice (two to eight), yet their gains are not optimal, due to their self-centered utilization of the advice. The findings suggest that in combining sets of three to nine opinions altogether (i.e., their prior opinions *and* the advice), participants selectively weight the opinions that are close to their own, while ignoring those that are distant from their own prior opinion. We call this egocentric trimming. This result contrasts with our normative analysis (based on the same data), which suggests that trimming is indeed a good strategy that could be used beneficially to improve accuracy, as long as it is conducted *objectively* rather than egocentrically (consensus-based trimming).

Aside from their practical implications for realistic decision making, these results carry a deeper theoretical message. A fundamental question in the literature in recent decades has been how adaptive or rational human behavior is, in light of research suggesting flaws in intuitive judgment and decision making. The process of giving and receiving advice could be viewed as an adaptive social decision-support system that helps individuals overcome their inherent limitations by proposing new alternatives, different frames, and disconfirming information.

The influence of advice: The process of revising opinions

How might people revise their opinions on the basis of advice from others? The task seems taxing, both cognitively and emotionally, as decision-makers need to decide how much weight to place on each opinion. This is especially difficult when the advisory opinions contradict each other or at odds with the decision-maker's own initial opinion. Two central concepts in this work are *egocentric judgment* and *satisficing*. We review their roles in advice-taking.

Egocentric judgment

Self-centered judgments are common in social settings (e.g., Chambers & Windschitl, 2004; Dunning & Hayes, 1996). Our previous findings suggest that people tend to be *egocentric* in revising their opinions. In particular, they tend to discount advice and favor their own opinion (Yaniv & Kleinberger, 2000; Yaniv, 2004a). This self/other effect has been observed in experiments using a "decide–advice–revise" paradigm where respondents form initial opinions and then revise them on the basis of one piece of advice. Consistent findings have been reported by others. In a cue-learning study by Harvey and Fischer (1997), respondents shifted their estimates about 20–30% towards the advisor's estimates. In another study by Lim and O'Connor (1995), judges weighted their own forecasts more heavily than advisory (statistical) forecasts. Sorkin, Hayes, and West (2001) reported a related result based on a group signal-detection task. Finally, in a study involving the control of a simulated system, Gardner and Berry (1995, Experiment 1) report that participants ignored useful advice when it was given to them as an option.

Yaniv (2004b; Yaniv and Kleinberger, 2000) suggested the following explanation for this finding. From an *external* (objective) point of view, a respondent's initial opinion and the advisor's opinion are on equal footing. However, from the decision-maker's *internal* (subjective) point of view, his or her own opinion and those of others are *not* on equal footing. Individuals are privy to their own thoughts, but not to those of others. They have less access to evidence supporting the advisor's view. The egocentric weighting of advice then results from the nature of the support the decision-maker can recruit for her own opinion versus the advice. Hence, other things being equal, decision-makers tend to discount advice.

A second egocentric effect is the *distance effect*, according to which individuals give less weight to advice the further it is from their initial opinion (Yaniv, 2004b). This finding is reminiscent of similar findings in the attitude-change literature. A basic tenet of all consistency theories of attitude change is that individuals seek to resolve discrepancies among their beliefs. Such theories predict that attitude change should decline with distance (Aronson, Turner, & Carlsmith, 1963; Sherif & Hovland, 1961). Bochner and Insko (1966) presented a persuasive message advocating that people get some specific number of hours of sleep per night (where the number ranged in various conditions from 8 to 0 h). They found that as the advocated number of hours of sleep decreased (the discrepancy increased), the magnitude of attitude change decreased (assuming that change is expressed as a fraction of the distance between the initial attitude and the message). As the message became more extreme, people generated more counter arguments and tended to disparage the source.

This distance effect was seen also in studies of stereotype change (Kunda & Oleson, 1997), and conceptualized in terms of assimilation and contrast processes (Sherif & Hovland, 1961; Wegener, Petty, Detweiler-Bedell, & Jarvis, 2001). While a slightly deviant opinion can be assimilated and thus cause a shift in one's attitude, a highly discrepant one accentuates the contrast; it has a reduced effect, since it falls outside the person's "latitude of acceptance" (Sherif & Hovland, 1961). Davis et al. (1997) also incorporated this idea into their social decision schemes. Their models describe how the opinions of groups (e.g., committees, juries) are aggregated during discussion to establish the group's consensual judgment. In their models, a discrepant opinion's impact on group decision quickly declines as the discrepancy increases.

Processing opinions from multiple sources

The findings reviewed so far demonstrate the egocentric effects in the processing of a *single* piece of advice. Here we consider the egocentric processing of *multiple* pieces of advice. Investigating the processing of a number of pieces of advice is important for practical and substantive reasons. First, the number of opinions in real-life decisions varies. Patients facing non-trivial health problems often seek a second and even a third expert opinion. Editors typically solicit the opinions of two or three reviewers to make publication decisions; universities seek perhaps three to six recommendation letters prior to making job offers or making tenure decisions. Is it only that resource constraints (time, effort, money) limit the number of opinions searched prior to making decisions? Or does experience tell decision-makers that polling a few opinions may suffice, on average, to exhaust most of the information that could be possibly obtained? Perhaps both factors affect the number of opinions people seek. We explore systematically how the number of opinions presented to decision-makers (two to eight) affects how people use multiple sources (i.e., processing difficulty) and how much they benefit from them (i.e., marginal gains).

As the number of pieces of advice increases, so does the complexity of the integration. Task complexity generally leads people to rely more on heuristic shortcuts and to seek satisficing rather than optimal outcomes. For instance, when faced with a complex multidimensional choice (e.g., shopping) involving a large number of alternatives (e.g., brand names) and attributes that describe each alternative (e.g., price, quality, delivery), decision-makers turn to heuristics that reduce the amount of information considered (Payne, 1976). In the present context, the need to integrate across conflicting opinions may lead people to employ satisficing rules and engage in data reduction. Finally, some important parameters in studies of conformity and majority influence on individuals are the size of the group and the size of the majority (e.g., Brown, 2000, Chap. 4). By varying here the numbers and configuration of advisory opinions we could explore the patterns of influence of advice.

The benefit of combining advisory opinions

How beneficial is the use of multiple advisory opinions? One might wonder whether non-expert advice is useful at all. In fact, advisors need not be smarter or more knowledgeable than the receiver of the advice to be valuable. For example, in a study involving estimation, participants reduced their initial error by about 20% by considering just one opinion of a fellow student drawn at random from a pool (Yaniv, 2004b).

There is ample evidence that averaging the opinions of several individuals increases accuracy. For example, a study of the accuracy of inflation forecasts found that averaging the opinions of several forecasters were superior to selecting the judgment of any of the individuals (Zarnowitz, 1984). While an individual forecaster might have outperformed the average on occasion, none did so consistently. Such results have been demonstrated in diverse domains, ranging from perceptual estimations of line lengths to business forecasts, and are an important motivation for research on combining estimates (Armstrong, 2001; Ashton & Ashton, 1985; Larrick & Soll, 2006; Libby & Blashfield, 1978; Surowiecki, 2004; Winkler & Poses, 1993; Yaniv, 1997; Yaniv & Hogarth, 1993).

A number of formal models provide a theoretical basis for understanding when and how combining estimates improves accuracy (whether accuracy is measured in terms of mean absolute error or judgment-criterion correlation). These include binary-choice models based on the Condorcet jury theorem (majority rules/binary issues) and group signal-detection theory (Sorkin et al., 2001), models for combining subjective probabilities from multiple judges (Budescu, Rantilla, Yu, & Karelitz, 2003; Wallsten, Budescu, Erev, & Diederich, 1997), and models for combining point forecasts (Clemen, 1989; Hogarth, 1978). In the case of quantitative judgments, a brief outline can show how the use of advice might improve judgmental accuracy. According to the Thurstonian view, a subjective forecast about an objective event is the sum of three components: The "truth," a systematic bias, and random error. Statistical principles guarantee that forecasts formed by averaging several sources have lower variability (random error) than the individual opinions. The combined forecasts are expected to converge about the truth if the systematic bias is zero or fairly small (e.g., Einhorn, Hogarth, & Klempner, 1977).

Overview of the research

A key feature of the present research is its dual emphasis on descriptive and normative issues. First, we ask what policies or rules people use for revising their opinions when presented with multiple advice. In other words, how do advisory opinions influence people's final opinions? Second, we ask what policies or rules improve decision accuracy.

The experiments, which were conducted on a computer due to their interactive nature, 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 several advisory estimates drawn from a large pool of estimates made by other students. The respondents were then asked to provide their estimates once again. They were free to use the advisory opinions as they wished. In the first experiment, the number of advisory opinions presented varied from two to eight in different conditions and the advice were selected on-line at random by the computer from appropriate pools of estimates.

What rules might people use in revising their opinions? Conceivably, there is an infinite number of potential rules that decision-makers could use for aggregation. It is practically impossible to test or even enumerate any great number of them. We therefore focused on relatively simple heuristic revision rules. We assumed that (a) individuals seek to produce the most accurate estimates they can (adaptiveness), but (b) they rely on simple heuristic rules when dealing with conflicting opinions (satisficing), and (c) their own perspective plays an important role in the revision process (egocentrism).

We considered an array of revision rules, among them one that assign equal weights to all opinions and others that discount some of the opinions. Some heuristics reduce dissonance through data reduction; they simplify the combination of opinions by trimming the set of opinions. Two types of trimming were considered. With egocentric trimming, the one (or two) opinions furthest from the decision-maker's own opinion are dropped from consideration. With consensus-driven trimming, the opinions furthest from the group's consensus are dropped. Thus extremity of opinion is defined subjectively (egocentrically) in the former and objectively in the latter case.

The egocentric trimming rule was designed to evaluate the hypothesis that judges weight distant (incompatible) opinions egocentrically. The consensus trimming rule was needed as a comparison with egocentric trimming. More importantly, there is some evidence that consensus trimming improves accuracy above equal weighting (Yaniv, 1997). The discussion presents some theoretical arguments justifying such trimming. Our descriptive analyses of how people revise their opinions are accompanied by a parallel normative data analysis designed to evaluate the adaptive value or success of each revision policy.

In the second experiment, the advisory opinions were not sampled from realistic pools, but were created artificially by design. The artificial profiles of near and far advice enabled us to conduct a series of linear regression analyses and compare weights (coefficients) for the two kinds of advice and also compute weight indices akin to those used in earlier studies (Yaniv & Kleinberger, 2000). In sum, the experiments used a variety of data-analytical approaches in an attempt to reach converging conclusions.

Experiment 1

This experiment investigated how people integrate their prior opinion with those of the advisors. The number of advisory opinions was two, four, or eight. The amount of advice were manipulated among participants. Two important notes are in order. First, our respondents received a bonus for making accurate judgments, so their decisions were consequential. We paid a bonus for each final estimate with a lower than average error, so it was in the respondents' interest to consider the advice carefully and make the best use of it, in whatever way they thought was appropriate. Second, a major advantage of the method of Experiment 1 is the representative sampling (Brunswik, 1955) of advice from pools of actual estimates made by other undergraduate students whom they might also have consulted in a natural situation.

Method

The experimental procedure was conducted individually on personal computers. Twenty-four 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 and asked to type in their best estimate for each one via the computer keyboard; in addition, they were asked to give lower and upper boundaries such that the true answer 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 which they would be presented with the same set of questions again. Now, however, each question would be presented along with the respondent's own estimate (from the initial phase) and several advisory estimates made by others. The respondents would then be asked to give a second, possibly revised, estimate for the question. No online feedback was given on the accuracy of their own or the advisors' opinions (in particular, the correct answers were never shown). The respondents were told they would get a bonus at the end of the study, depending on their overall accuracy (see below).

In one condition ($N = 55$), respondents were given two advisory estimates on each trial. In another condition ($N = 61$), they were given four opinions on each trial. In a third condition ($N = 54$), they were given eight opinions on each trial. The advisors' estimates were randomly drawn by the computer from a pool of 90 estimates collected in an earlier study in which respondents had been instructed merely to provide the best estimate for each question. For each question, new advisors were sampled at random. Thus the advisors varied from one question to the next, with labels such as #37, #81, and #15 used to indicate that the estimates came from different individuals in each trial. By adhering to representative sampling of estimates from pools of data, we insured that the dispersion of the estimates and their errors corresponded to those that might have been encountered in reality by our respondents when seeking answers to such questions among their peers—undergraduate students.

The respondents were undergraduate students who participated either as part of their course requirements or for a flat fee of $4. They were all told that they would receive a bonus based on the accuracy of their estimates. In particular, they would receive $0.25 as a bonus for each estimate with a better than average accuracy score. Altogether they could collect up to $6 in bonus payments. Thus it was in their interest to pay attention to the advisory estimates and make the best use of them. The bonus was based on the final estimates (i.e., second phase).

Table 1
Sample Question and Outline of the General Procedure

Phase 1 (series of 24 questions):	
In what year was the Suez Canal first opened for use?	
Your best estimate ________	
low estimate ____ high estimate ____	
Phase 2 (same 24 questions repeated):	
In what year was the Suez Canal first opened for use?	
Your previous best estimate was	1890
The best estimate of advisor #33 was	1905
The best estimate of advisor #16 was	1830
Your final best estimate	____

Results

We carried out two sets of analyses. First, we evaluated the accuracy of the initial and revised estimates along with the accuracies of the estimates generated by each of the formal revision rules listed below. Second, we evaluated which revision rules best fit the participants' final estimates. In accord with our assumption that processing advice is both heuristic and egocentric, we included revision rules that were based on simple descriptive statistics and seemed easy to execute, as well as ones that gave greater weight to the respondent's own perspective. The rules differ systematically from one another, and thus permit comparisons aimed at unveiling which operations people might use.

We assessed the fit of the following revision rules: *Equal weighting* (*average*) is the simple average of all opinions (i.e., including one's own and the advice). *Discounting extremes* (*median*) is the median of all opinions. *Midrange* is the average of the highest and lowest opinions in the set. With *consensus-based trimming* (*type one*) the most extreme opinion is removed and the remaining opinions are simply averaged, while with *consensus-based trimming* (*type two*) the two most extreme (i.e., the highest and lowest) opinions are removed and the remaining ones are simply averaged. With the *egocentric trimming* (*type one*) the single advisory opinion that is furthest from the respondent's initial opinion is removed and the remaining opinions are averaged, while with the *egocentric trimming* (*type two*) the two furthest opinions from the respondent's initial opinion are removed, and the remaining opinions are simply averaged. We analyzed the fit of the various rules and used significance testing to assess the differences among them.

The benefit of multiple advice

Table 2 shows the accuracy of the intuitive estimates and the estimates obtained by the formal revision policies. The results for each advice condition are shown in a different row. The actual success of intuitive revision is seen in the accuracies of the *initial* and the *final estimates*, as shown in the leftmost two

Table 2
Analysis of the benefit of advice: accuracy (mean absolute error, in years) of intuitive judgments and revision rules

Number of opinions	Stay with initial	Final	Median	Mean	Consensus trim 1	Consensus trim 2	Egocentric trim 1	Egocentric trim 2	Midrange
Self + 2	70.4	51.7	**<u>46.1</u>**	**46.9**	48.9	—	55.3	—	50.1
Self + 4	69.6	50.2	**41.1**	42.8	41.2	**<u>40.4</u>**	44.6	50.0	50.6
Self + 8	71.5	47.8	**<u>36.9</u>**	40.7	**38.0**	38.8	40.2	41.3	52.0

Note. The numbers are the mean absolute errors resulting from estimating the truth according to each of the schemes.
The best fit appears underlined and bold; the second best appears in bold.

columns of Table 2. The respondents' estimates improved due to the advice. The magnitude of improvement (i.e., reduction in mean absolute error) was roughly 27% after getting two opinions, 28% after getting four opinions, and 33% after getting eight opinions. The gains were substantial, and increased monotonically as a function of the number of opinions, but the *marginal* gains diminished rapidly as the number of opinions increased. The other columns in Table 2 show the accuracies of the revision rules. The best rules in each row are highlighted. Overall the median and consensus policies fared best.

In subsequent analyses, we compared the performance of the various policies using significance testing. Table 3 presents the sign test results for all pairwise comparisons between the accuracies of all the revision policies. The rules were listed in decreasing order of accuracy (i.e., the mean absolute error increases from left to right). Each entry indicates the number of times that a given *column* rule outperformed the corresponding *row* rule. As an illustration, Table 2 shows that the mean absolute error (in years) was reduced from 70.4 (initial-self estimate) to 51.7 (final-self estimate). Table 3 shows that final estimates were more accurate than the initial ones for 51 out of 55 participants, a significant difference, by the sign test, $N^{+} = 51$ (out of $N = 55$), $N^{-} = 4$, N-ties = 0, $p < .01$. Ties were generally rare and are indicated in the table whenever they occurred.

Below we summarize the significant sign test results in a schematic graph. Policies are listed from left to right in increasing order of accuracy. Policies that were not significantly different from each other at $p < .01$ are joined by an underline. For example, in condition 1, the median was better than all other policies, except for the average. In condition 2, both types of consensus-based trimming and the median policy did not differ from each other; but consensus trim-2 and the median were more accurate than the remaining policies. (This representation is not always easy to draw and depends on the "orderliness" of the significance of the pairwise comparisons.)

```
Condition 1: Self + 2 other opinions

Initial   Egocentric  Midrange   Final        Consensus    Average    Median
          Trim 1                              Trim 1
                      _____________===============
                                              ++++++++++++++++
                                                           --------------------------

Condition 2: Self + 4 other opinions

Initial   Midrange    Final    Egocentric  Average   Egocentric  Consensus     Median      Consensus
                               Trim 2                Trim 1      Trim 1                    Trim 2
     ____________________________              +++++++++++++++++++++++++
                                                                 ------------------------------------------------

Condition 3:Self + 8 other opinions

Initial   Midrange  Final   Average    Egocentric   Egocentric   Consensus    Consensus     Median
                                         Trim 2       Trim 1        Trim 2     Trim 1
                                                                               ----------------------
                    ____________         +++++++++++++++++++++++++
                            ^^^^^^^^^^^^^^^^^^^^^^^^^^^
           ============
```

The ordering of the rules was similar, though not identical, in all conditions. The data permit several conclusions. We find that the median rule is better than the average with five (self + 4) and nine (self + 8)

Table 3
The accuracy gains of intuitive revision and the various revision rules: sign test pairwise comparisons (entries show the number of times the column rule outperforms the row rule and the number of ties (=); *5% ; **1%)

	Median	Average	Consensus trim 1	Midrange	Self final	Egocentric trim 1		
Condition 1: self 2			$N = 55$					
Self initial	41**	47**	46**	47**	51**	46**		
Egocentric trim 1	44**	44**	38**	41**	39**			
Self final	40**	38**	30	30				
Midrange	41**	51**	36*					
Consensus trim 1	45**	30						
Average	34;1 =							
	Consensus trim 2	Median	Consensus trim 1	Egocentric trim 1	Average	Final	Egocentric trim 2	Midrange
Condition 2: self + 4			$N = 61$					
Self initial	54**	53**	52**	58**	53**	53**	58**	47**
Midrange	58**	55**	55**	51**	59**	35	32	
Egocentric trim 2	46**	47**	46**	52**	44**	34		
Self final	47**	47**	47**	45**	44**			
Average	47**	41; 1 = **	38	39*				
Egocentric trim 1	46**	46**	39*					
Consensus trim 1	38	32						
Median	35;1 =							
	Median	Consensus trim 1	Consensus trim 2	Egocentric trim 1	Egocentric trim 2	Average	Self final	Midrange
Condition 3: self + 8			$N = 54$					
Self initial	51**	50**	49**	51**	53**	48**	53**	40**
Midrange	52**	53**	54**	53**	52**	54**	34	
Self final	42**	41**	40**	42**	43**	36*		
Average	45**	44**	50**	32	30			
Egocentric trim 2	40**	37**	34	35*				
Egocentric trim 1	44**	38**	32					
Consensus trim 2	40**	37**						
Consensus trim 1	34							

Table 4
Which rule best explains the influence of advice? Fits of revision rules (mean absolute deviation, in years)

Number of opinions	Stay with initial	Final	Median	Mean	Consensus trim 1	Consensus trim 2	Egocentric trim 1	Egocentric trim 2	Midrange
Self + 2	36.2	—	**23.9**	31.5	29.5	—	**27.9**	—	36.5
Self + 4	38.3	—	29.3	34.2	32.8	31.1	**27.9**	**27.5**	42.6
Self + 8	38.2	—	33.0	35.5	34.8	34.6	**31.1**	**29.0**	45.6

Note. The numbers are the mean absolute deviations indicating the fit of each rule to the final opinion. Lower numbers indicate better fit of the rule. The best fit appears underlined and bold; the second best appears in bold.

opinions, and is as good as the average with three (self + 2) opinions. The two consensus heuristics performed well with five to nine opinions, but not with three opinions, presumably since such a small number of opinions does not permit useful removal. One conclusion is, briefly, that operations that attenuate the influence of extreme opinions tend to increase accuracy. Overall the median (which effectively attenuates the influence of extremes) and the consensus heuristic (which literally removes opinions) performed best. A second conclusion is that individuals were able to benefit from the advice, but failed to extract all the information contained in it. Intuitive final estimates were less accurate than median, consensus trimming and averaging.

The influence of advice

What do people actually do? How do they open themselves to the influence of advice? Each revision rule represents a pattern of influence of multiple advice on respondents' final estimates. To assess this influence, we calculated the fit of each rule with the final estimates, defined as the distance between the rule's prediction and the actual final estimate. Table 4 shows the global fit of each rule (i.e., the mean distance) in each advice condition (smaller numbers indicate better fit). Note first the rule labeled "initial self"—the strategy of staying with the initial estimate. This rule ignores all advice, thus technically representing the highest level of egocentrism. We found, in line with earlier findings, that judges adhered to their initial estimates in 38, 39, and 40% of the cases, in conditions 1–3,

Table 5
Which rule best explains the influence of advice? Pairwise comparisons (sign tests) of rules (entries show the number of times the column rule outperforms the row rule; *5%; **1%)

	Median	Egocentric trim 1	Consensus trim 1	Average	Self initial
Condition 1: self + 2			$N = 55$		
Midrange	51**	42**	41**	54**	28
Self initial	39**	42**	30	31	
Average	51**	38**	36*		
Consensus trim 1	47**	34			
Egocentric trim 1	31				

	Egocentric trim 1	Egocentric trim 2	Median	Consensus trim 2	Consensus trim 1	Average	Self initial
Condition 2: self + 4			$N = 61$				
Midrange	61**	57**	55**	57**	49**	58**	37
Self initial	39*	42**	35	34	33	31	
Average	54**	49**	49**	51**	37		
Consensus trim 1	48**	42**	47**	41**			
Consensus trim 2	49**	38	42**				
Median	38	35					
Egocentric trim 2	32						

	Egocentric trim 2	Egocentric trim 1	Self initial	Median	Consensus trim 2	Consensus trim 1	Average
Condition 3: self + 8			$N = 54$				
Midrange	53**	51**	39**	48**	48**	46**	51**
Average	48**	50**	30	44**	38**	33	
Consensus trim 1	46**	45**	29	41**	33		
Consensus trim 2	47**	48**	30	39**			
Median	40**	37**	28				
Self initial	30	29					
Egocentric trim 1	43**						

respectively. (They were correct less than 5% of the time in their initial estimates.) The two other egocentric rules ignore only the most distant (one or two) pieces of advice. The consensus rules, average, median, and midrange provide important benchmarks and are interesting because we have information about their accuracy.

The two best-fitting rules are highlighted in each row. In condition 1 (self + 2), the median emerged as the best-fitting heuristic, followed by egocentric trimming. With more opinions (self + 4, self + 8), the best-fitting heuristics involved egocentric removal of distant opinions. Of importance is the finding that the fit of egocentric trimming was better than that of objective trimming and equal weighting. The midrange was the worst fit. These results provide meaningful comparisons. The difference between the processes of egocentric trimming and averaging involves just one operation, namely, the deletion of one or two opinions selectively. The difference between the processes of consensus trimming and egocentric trimming is the selection of the opinion(s) to be deleted prior to averaging. Finally, the midrange depends only on the extremes, in contrast with the median, which seeks the center and attenuates the effect of the extremes.

We conducted sign tests to compare the fits of the various rules. The results of all pairwise comparisons are shown in Table 5. For illustration, the comparison between the median and the midrange yielded results as follows: $N^{+} = 51$ (out of $N = 55$), namely, $N^{-} = 4$, $N^{0} = 0$. (In general, ties are rare and are indicated when they occurred.) As can be seen, in condition 1 the median had a better fit than the other rules, except for egocentric trimming. The comparisons by and large support the conclusion that the egocentric heuristics provided better fits than the other schemes.

Below we present a schematic summary of the pairwise comparisons shown in Table 5. The heuristics are shown in increasing order of fit from left to right, according to Table 4. Two heuristics were considered different if they differed at the $p < .01$ level of significance; thus heuristics that did not differ at this level of significance were joined by an underline. This representation provides only an approximate summary of the results shown in Table 5. The main exception to the linear ordering of the heuristics occurred in condition 3, where the fits of the egocentric rules were not significantly better than those of the initial self (although the numbers shown in Table 4 suggest that the egocentric policies might have significantly better fit). The overarching conclusion from these analyses is that judges use advice egocentrically, either by egocentric trimming or by the most egocentric strategy of all—staying with one's initial opinion. It is worth noting that the egocentric heuristics provided better fits of judges' estimates, although they were inferior to the consensus and median heuristics in terms of accuracy.

Condition 1: Self + 2 other opinions

```
Midrange          Initial    Average     Consensus   Egocentric    Median
                  Self                   Trim 1      Trim 1
                  ________________       ++++++++++++++++
=====================                        --------------------------
```

Condition 2: self +4 other opinions

```
Midrange   Initial    Average    Consensus  Consensus  Median     Egocentric  Egocentric
           Self                  Trim 1     Trim 2                Trim 1      Trim 2
++++++++++++++                                         ________________________________
           --------------------------
           ^^^^^^                           ^^^^^^^^^
           =====                 ======
```

Condition 3: self +8 other opinions

```
Midrange   Initial    Average    Consensus  Consensus  Median     Egocentric  Egocentric
           Self                  Trim 1     Trim 2                Trim 1      Trim 2
           ++++++++++++++++++++++++++
           _____                 ---------------------------      ________
           ====                                        =====
```

Conclusions

A major advantage of sampling ecological advice, rather than creating the advice by design, is that it permits us to conduct descriptive and normative analyses—that is, to contrast the actual use of multiple advice and the strategies that appear to increase the benefits that could be reaped by using advice—all within the same experiment. Several findings are notable. First, individuals use advice egocentrically, as the fits of the egocentric trimming rules suggest. Second, their accuracy gains are substantial. Third and more striking is the finding that the intuitively revised estimates are about as accurate as the worst (midrange) method (Table 2), and significantly worse than the egocentric trim strategies, the median, and the consensus trim strategies. In the final discussion, we consider these findings and present principled arguments for why consensus trimming boosts accuracy most.

Experiment 2

This experiment further investigated the use of multiple advice. Unlike the previous experiment, which involved representative samples of advice (i.e., random draws from a pool), here the advice was constructed. The advice was created online for each participant, depending on the estimates he or she gave in the first phase. On each trial, two advisory opinions were created, one near and one far from the participant's initial opinion.

We investigated how near and far advice influence one's final opinion (accuracy was of less importance since the advice were manipulated rather than sampled). Using different data-analytical methods, we assessed, first, the utilization of each type of advice (near vs far); second, how participants' prior knowledge affects their utilization of advice; and third, the effects of the quality of the advice on how it is used. The findings allow us to assess the revision rules and the theoretical explanations of egocentrism in opinion revision.

Method

Procedure

The procedure included two phases, as in Experiment 1. After producing initial estimates, respondents ($N = 75$) received the same questions again along with the advice and were instructed to provide their final estimates.

For each question two pieces of advice were generated online by the computer at the end of phase 1 (i.e., after the initial estimates were entered). The pieces of advice (near and far estimates) were a function of the participant's estimate in phase 1; they were generated by either adding or subtracting certain factors to or from the initial estimates. The factors for a given question depended on the actual spread of the pool of answers estimating that question. For each question, we calculated the inter-quartile range (IQR) of a pool of 70 estimates (collected in earlier studies). The factors added to (or subtracted from) the initial self opinion were a function of IQR; specifically, for the near advice the factor was 0.3∗IQR and for the far advice, 0.9∗IQR. Using IQR in creating the factors lent credibility to the advisory opinions by ensuring that they were well within the natural spread of estimates for each question. In addition, the different questions had different spreads of answers (IQR) and hence different factors for each question.

There were four conditions, each involving a different configuration of the near and far advice vis-à-vis the truth. In condition 1, both pieces of advice always pointed *away from the truth* (technically, when the initial opinion was below the truth the relevant factor was subtracted from it, and when the initial opinion was above the truth the factor was added to it). In condition 2, both pieces of advice pointed *towards the truth* (i.e., when the initial opinion was below the truth the relevant factor was added to it, and when the initial opinion was above the truth the factor was subtracted). In condition 3, the far advice pointed away from the truth, while the near advice pointed towards the truth. In condition 4, the far advice pointed towards the truth, while the near advice pointed away from the truth.

There were a total of 28 trials. We used the same 24 questions as in Experiment 1, plus four new ones. Seven questions were randomly assigned to each of the four advice configurations, and trials in all four conditions were presented in a new random order for each participant. Thus advice configuration was a within-subject factor with four levels. The instructions were the same as in Experiment 1. Nothing was said to the participants about how the estimates were created. The participants were told that they could earn up to $7 as a bonus for accuracy. Hence it was in their interest to consider their answers carefully and make the best use of the advice provided.

Results

We analyzed the rules that participants used in revising their prior opinions as a function of participants' knowledge. Prior to the analyses participants were divided into high- and low-knowledge groups according to the accuracy of their estimates in the *first* phase. Using the median-split method, high-knowledge participants had mean absolute errors below the median and the low-knowledge ones had mean errors above the median.

The first set of analyses involved three variants of linear regression, a method that has been commonly used in analyses of implicit weighting policies in multiple-cue judgment tasks and "lens" models (e.g., Einhorn, 1972). With this method, the final estimates are regressed on three predictors: initial estimate, near advice, and far advice, with the coefficients representing the weights assigned to each opinion.

The estimation of the regression coefficients tends to be unstable if the predictors are intercorrelated (a common concern in judgment analysis). Therefore, we report the results based on three alternative approaches. First, we performed a stepwise procedure for *each* participant separately, focusing only on the first predictor selected in each regression. Second, we performed a simultaneous regression (entering all variables at once) for *each* participant, looking at systematic differences among the coefficients obtained *across* all 75 individual regressions. Third, we report the utilization indices as recommended in the literature on judgment analysis (Cooksey, 1996, pp. 165–167) to circumvent the problem of intercorrelation.

Stepwise regression

For each participant, the final estimates were regressed on three variables (initial estimate, near advice, and far advice) using the SPSS forward stepwise procedure (F-to-enter significant at $p = .05$). We considered here only *which* predictor was selected *first*. The first predictor selected was the initial estimate (the self) in 71% of the regressions (53 of 75), the near advice in 28%, and the far advice in 1%. Similar results were seen in both knowledge groups. In the high-knowledge group ($n = 38$), the first predictor selected was the *self* in 71% of the regressions (27 of 38), and the *near advice* in 29%. In the low-knowledge group ($n = 37$), the first predictor was the *self* was in 65% of the regressions, *near advice* in 32%, and *far advice* in 3%. By limiting our frequency statistics to the first predictor selected only (thereby ignoring the information on the remaining predictors), we bypass the intercorrelation issue. This conservative approach suggests that, overall, the self was the best predictor of the final estimates (followed by the near advice).

Simultaneous regression

For each participant, the final estimates were regressed on three predictors: initial estimate, near advice, and far advice. Three predictor coefficients were obtained for each participant; their averages are shown in Table 6 (left). The high-knowledge participants placed the largest weight on the self, lower weight on the near

Table 6
Regressions predicting final estimates from initial estimates and advice

Participants' knowledge	Regression coefficients			R^2	Utilization coefficients		
	Self initial	Near advice	Far advice		Self initial	Near advice	Far advice
High ($n = 38$)	0.57	0.35	0.10	0.97	0.0177	0.0116	0.0069
Low ($n = 37$)	0.39	0.38	0.23	0.96	0.0128	0.0079	0.0161
Overall ($n = 75$)	0.48	0.37	0.16	0.96	0.0153	0.0098	0.0115

advice, and the lowest weight on the far advice. The low-knowledge participants placed roughly the same weights on the self and near advice, and lower weight on the far advice.

The coefficients were compared statistically within groups by sign tests. For the high-knowledge group ($N = 38$), there was a significant difference between self and near advice, $N^+ = 26$ out of 38 (no ties), $p < .05$, and between near and far advice, $N^+ = 31$ out of 38 (no ties) $p < .001$. For the low-knowledge group ($N = 37$), there was a significant difference only between near and far advice, $N^+ = 28$ out of 37 (no ties), $p < .005$. Finally, between-group comparisons showed that the low-knowledge group paid more attention to the far advice than did the high-knowledge group (Mann–Whitney test, $z = 4.53$, $p < .001$). In sum, the simultaneous regression is generally susceptible to problems due to multicollinearity (intercorrelations among predictors), in that the estimation of the regression coefficients tends to be highly variable, and hence, less stable. Despite this variability, the obtained pattern was consistent across the sample of 75 judges, and also consistent with the stepwise results, suggesting preference for the self and the near advice.

Utilization indices

Cooksey (1996, pp. 165–167) recommends the use of "utilization indices" when the predictors are intercorrelated. The utilization of the ith predictor is its unique contribution—what one loses in predictive power by deleting it from the regression equation. It is computed as the difference between the full model (including three predictors here) and the reduced model including all but the ith predictor. In other words, it is the squared semi-partial correlation of the ith predictor with the criterion. The utilization indices were computed for each participant and then averaged (Table 6, right).

The utilization indices were compared within and between the knowledge groups. The high-knowledge group had utilization indices that were ordered (self > near advice > far advice) like the regression coefficients; the three pairwise comparisons were significant at $p < .05$, by the one-tailed sign test (self vs near, near vs far, self vs far), $N^+ = 25$ out of 38 (N^+ was the same in all comparisons, no ties). The low-knowledge group utilized the far advice more than the near advice, $N^+ = 29$ out of 37 (no ties), $p < .05$, but the self vs far contrast was not significant, $N^+ = 29$, $p > .5$. Finally, Mann–Whitney comparisons *between* the knowledge groups showed that the low-knowledge group utilized the far advice more than the high-knowledge group ($z = 3.11$, $p < .05$) while the high-knowledge group utilized the self more than the low-knowledge group ($z = 2.99$, $p < .05$). In sum, the utilization analyses suggest that the low-knowledge group utilized the far advice more than was revealed by the previous analyses. The various regression approaches generally agreed, but not always. Cooksey (1996) notes that differences between the utilization indices and the regression coefficients are to be expected if the predictors are intercorrelated.

The weight index

The second analysis was meant to complement the regressions. We created an index of the location of the final estimate relative to the initial opinion and the advice. We defined the "weight index" = |f-aa|/|i-aa| where *i*, *f*, and *aa* stand for initial self opinion, final, and average advice, respectively. With this index, the final estimate is represented as a weighted combination of the *initial self estimate* and the *average of the two pieces of advice*. This weight measure, which is akin to those used earlier (Yaniv & Kleinberger, 2000; Yaniv, 2004b), enables us to compare behavior across the four experimental conditions.

As an illustration, suppose the initial estimate is 1850 and the average advice is 1880. The weight index takes a value of 1.0 if, in making the final estimate, the judge adheres completely to her initial estimate (1850); and 0 if the judge shifts completely to the average advice (1880). Intermediate weights between 0 and 1.0 indicate that the final estimate was in the range between the initial self and the average advice (1850–1880). A index value lower than 0 indicates that the judge shifted even beyond the average advice (e.g., above 1880), while values greater than 1.0 indicate that the judge shifted beyond the self (e.g., below 1850). Equal weighting involves placing weights of one-third on the self and two-thirds on the average advice.

The mean weight index—which can be interpreted as the participant's weight for her own initial opinion—is shown as a function of participants' knowledge in Table 7. A weight index of one-third (0.33) would be obtained if participants relied on their own opinion no more but also no less than they relied on advice. The actual mean weight (0.71) was significantly higher than 0.33, $t_{74} = 14.6$, $p < .001$, suggesting that participants relied more on their own opinion than on the advice. (This result is consistent with earlier ones with just one advisory opinion; Yaniv & Kleinberger, 2000). High-knowledge participants had higher weight indices than low-knowledge ones (0.83 vs 0.60), $t_{73} = 5.14$, $p < .001$. Thus the high-knowledge participants relied less on the (average) advice.

The mean weight index, shown in Table 8 as a function of advice configuration, suggest that participants

Table 7
Weight of own opinion as a function of participant's knowledge

	Participant's knowledge		Overall ($n = 75$)
	High ($n = 38$)	Low ($n = 37$)	
Mean	0.83	0.60	0.71
SD	0.34	0.52	0.45

Table 8
Weight of own opinion as a function of advice condition

	Advice condition: location of near and far advice relative to the truth			
	Both away from truth	Both toward truth	Near toward truth (far away)	Far toward truth (near away)
Mean	0.63	0.54	1.01	0.67
SD	0.30	0.30	0.47	0.55

had valid intuitions about the quality of the advice. Consider the two leftmost conditions. Participants correctly leaned less on their own opinion and shifted more towards the average advice in condition 2 (*both towards*) than in condition 1 (*both away*) (0.54 vs 0.63), $t_{74} = 2.53$, $p < .05$. Similarly, a comparison between the two rightmost conditions shows that the participants placed more weight on the average advice when it pointed in the direction of the truth (*far towards the truth*) than when it pointed away from the truth (*near towards truth*) (0.67 vs 1.01), $t_{74} = 3.71$, $p < .05$.

Accuracy gains and losses due to advice

Unlike Experiment 1, accuracy was not a focal issue here since the advice was manipulated. As might be expected, the accuracy of the final estimates improved or declined depending on the type of advice presented. We observed a 15% accuracy gain in condition 2 (*both towards*), and a 20% accuracy loss in condition 1 (*both away*). In the mixed conditions 3 and 4, the changes were minute: a 3% gain in condition 3 (*far towards the truth*) and a 1% loss in condition 4 (*near towards truth*). One-way analysis of variance on the gain/loss variable (absolute error of initial estimate minus absolute error of final estimate) found significant differences among the four conditions, $F_{3,296} = 76.9$, $p < .05$. Specifically, conditions 1 and 2 differed from each other (Tukey-HSD, $p < .05$); conditions 3 and 4 did not differ from each other, but they each differed from both condition 1 and condition 2.

Clearly, good advice helps decision-makers, while poor advice leads them astray. Gains are a function of the quality of the advice as much as of the revision rules that one uses. These findings seem to underscore the importance of representative sampling of advice in assessing accuracy gains.

Conclusions

According to the stepwise- and simultaneous-regression results, all participants used egocentric revision rules. The utilization analyses revealed slightly different results. Whereas the high-knowledge group also used egocentric rules, the low-knowledge group's utilization indices for *far advice* and *self* did not differ significantly from each other. While we do not know the reason for the difference, we note that according to Cooksey (1996), differences between the utilization indices and the regression coefficients may occur if the predictors are intercorrelated. Overall, these results indicate that participants' policies were sensitive to the *quality of their own knowledge*. The analyses based on the weight index suggest that participants were sensitive also to the *quality of the* (*average*) *advice*, giving it more weight in the conditions where it pointed towards, rather than away from, the truth. In sum, despite their general egocentric approach, participants were not oblivious to the quality of their own and others' opinions.

General discussion

Our experiments considered the process and consequences of using multiple advice. Specifically we investigated, first, how decision-makers integrate opinions from multiple sources of advice (process) and, second, whether and how much decision-makers gain from using such advice (consequences).

The benefit of advice: intuitive vs formal revision rules

Accuracy gains achieved by the judges

Exposure to a number of advisory opinions (either two, four, or eight in Experiment 1) helped participants improve their intuitive estimates dramatically—their accuracy gains were 27, 28, and 33% in the three conditions, respectively. But the marginal gains diminished quickly, since two opinions were enough to yield most of what could be gained by considering a larger sample of opinions. The seemingly puzzling result that additional opinions do not contribute much to accuracy has been observed in some earlier studies (e.g., Ashton & Ashton, 1985; Libby & Blashfield, 1978) and discussed theoretically (e.g., Hogarth, 1978; Johnson, Budescu, & Wallsten, 2001; Wallsten et al., 1997).

What could explain this phenomenon? Briefly, the accuracy gains resulting from aggregation are optimal if the advisors are independent. Gains of appreciable size may also occur where there are low or moderate positive correlations among experts (Johnson et al., 2001). However, the greater the *dependence* among the advisors, the lower the marginal gains of adding any one of them to the total. In many realistic situations—and our experiment is not an exception—some level of (statistical) dependence among advisors is to be expected. For instance, advisors may rely on similar information sources or have similar backgrounds (Soll, 1999);

moreover, in some real-life situations they may even consult one another, in which case additional advice is merely "more of the same." Along these lines, Sunstein (2003, chap. 9) considers the drawbacks of dependence among appellate judges in making the case for "why society needs dissent."

Apart from dependence, the presence of *systematic bias* further curtails the potential benefits of adding opinions. The greater the bias (e.g., systematic over- or under-estimation of the true value), the lower the gains from combining opinions. These structural considerations explain the fact that asymptotic accuracy levels are reached with very few advisory opinions.

Shortcomings of human revision

Although the potential accuracy gains from advice are limited in principle, individuals failed to exhaust even those that were possible. Normative analyses revealed that most formal revision policies were as good as or better than intuitive revision. The phenomenon that simple strategies (e.g., equal weighting) outperform intuitive judgments is conceptually analogous to classic findings from the study of linear (lens) models of judgment. Simple weighting policies that assign weights to the cues (predictors) consistently yield predictions that are more accurate than those produced by individual judges who have access to the same cues (Dawes, 1979; Einhorn, 1972; Einhorn & Hogarth, 1975).

The predominant explanation for this finding in the judgment literature is that the *inconsistent* application of a judge's weighting policy leads to inferior intuitive judgments. First, judges make random errors in applying their own judgment policy. Second, judges switch strategies or consider intricate interactions among the cues, rather than applying a consistent weighting of the cues. The net effect of both tendencies is poorer performance than with weighting policies that are consistently applied (Camerer & Johnson, 1991). The lesson from the judgment analysis literature is not negated here—our decision-makers were outperformed by equal weighting and trimming policies. Harvey, Harries, and Fischer (2000) also reached a similar conclusion. In their study participants estimated the monthly sales figures for an unknown product on the basis of forecasts made by four advisors. Accuracy feedback was also presented. Harvey et al. report that participants' global estimates were less accurate than a weighted average of the input advice.

When might ignoring advice be beneficial?

We obtained the intriguing finding that *the removal of extreme opinions* outperformed simple averaging. Clearly, an outlying opinion is not necessarily wrong. However, under certain conditions dissenting opinions are likely to be wrong, which would justify removing them (Harries et al., 2004; Yaniv, 1997). Assume a bell-shaped, thick-tailed distribution of opinions—that is, the prevalence of outlier opinions is larger than would be expected under the standard bell-shaped (normal) distribution. Under such conditions (assuming zero or small bias), an extreme opinion in a sample is particularly likely to be wrong. Given an underlying symmetric distribution with relatively thick tails, a *trimmed sample mean* is to be preferred to the raw sample mean as an estimator of the central tendency of the distribution (e.g., DeGroot, 1986, pp. 564–569; Wilcox, 1992; see also Streiner, 2000). Indeed, the distributions of human responses frequently have one or two thick tails, as Micceri (1989) has found. In sum, such statistical conditions warrant consensus-based trimming of extreme opinions.

To see the relevance of these arguments to small samples (e.g., of five or nine opinions), consider a *hypothetical binary* situation where 90% of the advisors' estimates are classified as near the true value, and the remaining 10% are far from the truth. In this scenario, far opinions impair accuracy fairly frequently. If a sample of five opinions is drawn at random from this population, then the chances of it including at least one far opinion is 35%. If the distribution is composed of 80% near opinions and 20% far opinions, then the chances of encountering at least one far opinion in a sample of five rises to almost 60%. The high likelihood of including far opinions in small samples (assuming thick tails) may explain why consensus trimming is such a powerful heuristic.

In sum, under some conditions removing dissenting opinions might increase accuracy. Removing opinions that are distant from the *self*, rather than from the *consensus*, raises different issues, however. In principle, by ignoring discrepant opinions that challenge them, people reduce their ability to learn and update their opinions. It has been shown how selective incorporation of evidence leads to perseverance of attitudes (Lord, Ross, & Lepper, 1979). In the same manner, egocentrically trimming discrepant opinions results in conservative revision of opinions. Our normative results indeed show that egocentric trimming is inferior to consensus trimming.

The influence of advice: intuitive revision rules

We assessed the fit of simple heuristic rules in Experiment 1 to uncover some of the mental operations that people use in processing samples of opinions. This is important since, as we suggested above, the benefits of using advice depend on *how* it is processed. While we did not locate a single rule that characterizes behavior on every trial (in fact, there might not be any), our data do provide evidence regarding people's approaches to the task. First, the revision process was highly egocentric, as several types of evidence indicate. In both experiments, the participants were correct less than 5% of the time in their initial estimates, yet they adhered to them in roughly 35–40% of the trials, changing their opinions only in the remaining 60–65%. Then, among the revision heuristics considered in Experiment 1, egocentric trimming provided the best approximation of the influence of 4 or 8 pieces of advice on participants' final estimates.

While Experiment 1 used representative (ecological) samples of advice, namely, opinions that were randomly drawn from pools of estimates, Experiment 2 used artificial (near and far) advice that was created systematically as a function of the participant's initial estimate. Several analyses suggest that decision-makers' revision policy here was egocentric, giving greater weight to confirming (i.e., near) advice than to disconfirming (i.e., distant) advice; at the same time, their revision policy was also sensitive to how much they knew to begin with. In Experiment 2, we analyzed high- and low-knowledge participants separately. The high-knowledge participants clearly placed more weight on their own opinion and on near advice than on far advice (stepwise and simultaneous regressions; utilization indices). The low-knowledge participants considered the advice more than did the high-knowledge participants (regression); the utilization indices showed that they generally placed more weight on the far advice than on near advice. Considering all the evidence, participants were sensitive to their own knowledge, showing greater utilization of the advice when they were actually less knowledgeable.

Our research adds to a body of results on the use of others' opinions, most of which involves *combining* opinions rather than *revision* of one's own prior opinion. In a recent study, participants combined a set of forecasts of the next day's temperature for each of a series of places in the UK (Harries et al., 2004). The actual places were not identified by name so the participants could not have any specific prior opinion. It was found that participants relied heavily on consensus trimming (egocentric trimming was obviously not an available strategy).

Budescu et al. (2003) suggested that judges aggregate opinions by a weighted average rule rather than a median rule. Two important aspects distinguish our setup from theirs. First, their studies involved combining probabilities (i.e., values drawn from a uniform distribution) whereas our setup involved estimates drawn from a distribution that had center plus "long tails" (i.e., numerous extreme estimates), thereby increasing the appropriateness of trimming operations. Second, in their paradigm participants *combined* others' opinions, whereas in ours participants *revised* their own initial opinions. Both of these aspects are likely to increase the tendency to discount distant opinions.

An account of egocentric judgment

How are we to explain the egocentrism in processing advice sets? First, selectively ignoring opinions that are inconsistent with one's own clearly simplifies cognitive processing (Harries et al., 2004; Yaniv, 1997). A second and a more profound rationale is based on the assumption that people take a subjectivist approach to belief updating (Hogarth & Einhorn, 1992). Thus they weight a new opinion as a function of their confidence in their *own* opinion and in the advice. People naturally have more access to the reasons underlying their personal opinion than to those underlying the advisors' opinions, hence they weight their own opinion more. Egocentric discounting results from judges' differential access to the evidence underlying each opinion.

The contrast between internally generated information and external sources has been pointed out in previous work as well. Koehler (1994) suggested that individuals assess the plausibility of a focal hypothesis differently, depending on whether they themselves selected it (in a binary-choice task) or the selection was made by someone else. More elaborate evidence (e.g., pros and cons) is retrieved for a self-generated opinion than for an opinion provided by someone else (see also the cueing effect in Ronis & Yates, 1987).

The observation that people have more access to the reasons underlying their personal opinion than to those underlying the advisors' opinions also explains other phenomena. The reasons underlying their personal opinion are more consistent with near advice than with far advice, hence they weight the former more heavily. This explanation is akin to other theories of attitude change (Lord et al., 1979) which posit assimilation and contrast processes; near advice is presumably assimilated, whereas far advice accentuates the contrast; advice that falls outside the "latitude of acceptance" is discredited and hence ignored (Sherif & Hovland, 1961).

Finally, disagreement among advisors is likely to boost egocentrism. Studies of multiple-cue judgments have shown that judges who are presented with inconsistent cues tend to place less weight on *all* of them (Slovic, 1966; Sniezek & Buckley, 1995). Our studies involved naïve advice, but the results may generalize to expert advice as well, whenever experts differ widely. For example, it is not uncommon for expert reviewers to present widely divergent views on an article submitted for publication. A vivid real-life example of expert disagreement appears in a recent *Nature* report entitled "Rival monsoon forecasts are banned" (Jayaraman, 2005). Since the Indian economy relies heavily on farming, accurate weather forecasts are critical. A ban on internet posting of "dramatically different" meteorological forecasts was called for, according to the Indian government, "to stop the public being confused by conflicting forecasts" (p. 161). Similar cases are not uncommon in the public arena. An open debate among experts on a given topic—say, about health risks associated with certain types of food, diet, or a new hazardous technology—decreases the public's trust in all expert opinions alike (Slovic, 1993). Disagreement among advisors presumably contributed to the egocentric effect in our experiments.

Cognitive vs motivational determinants of egocentrism

Whereas cognitive processes figure prominently in our account of the egocentric updating of opinions, motivational factors could conceivably also play a role in this process. For example, individuals are motivated to maintain consistency in their estimates; they may also be motivated to consider their own opinion superior to others' opinions in order to maintain their self-esteem. Cognition and motivation suggest different, though complementary, perspectives on egocentrism. In trying to assess the merits of each perspective, we begin by highlighting the theoretical distinction between the motivation to arrive at an *accurate* conclusion and the motivation to arrive at a particular, *directional* conclusion (Kunda, 1990). The first arises when individuals are rewarded for accuracy, whereas the second arises when individuals' self-esteem is hurt or threatened (Kruglanski & Freund, 1983).

The motivation to arrive at a favorable, self-serving conclusion could bias memory search, social judgment, and the evaluation of evidence (Pyszczynski & Greenberg, 1987). A number of studies have shown biased updating of prior opinions. In one study, coffee drinkers who were confronted with unfavorable evidence regarding coffee consumption tended to play down the information more than non-drinkers (Kunda, 1990). In another study, opponents and proponents of the death penalty systematically pointed out faults in evidence supporting the views that were opposite to their own (Lord et al., 1979). A second, related self-serving bias is one's tendency to commit to past decisions (Cialdini, 1993, Chap. 3). People seek to maintain consistency in their actions and thoughts and also wish to be viewed as consistent in social settings. Consequently they sometimes hold on to their past decisions and persist even in a failing course of action (Brockner, Rubin, & Lang, 1981). Finally, another self-serving bias that pervades interpersonal comparisons is the better-than-average effect. People tend to believe, for example, that they have less chance than others of experiencing negative life events, such as road accidents or strokes, or that they rank higher than others on various abilities and attributes, such as driving ability and social skills (Chambers & Windschitl, 2004; Kruger, 1999). How likely is it that effects of the sort reviewed above operated in our experimental setting?

The antecedents of commitment are often the high cost of being inconsistent, the need to justify decisions to others, having to admit past mistakes and the wish to save face (e.g., Brockner et al., 1981). In a similar vein, the antecedents of biased assimilation of evidence are threats to one's ego and one's cherished values. These antecedents were not present in our studies, however. Our participants made their judgments in a private setting (on a computer); they were not asked to justify their estimates and the questions did not involve any of their deeply held values. On the other hand, they received bonuses for accuracy, which presumably increased their willingness to revise their prior opinions. Indeed, there is evidence that the accuracy motivation induces effective information processing (Kruglanski & Freund, 1983).

Therefore, we believe that the accuracy motivation was stronger than the self-enhancement motivation. Indeed, our results show that less knowledgeable participants changed their opinions (based on advice) more than more knowledgeable ones (Experiment 2); also, the weight indices suggest that participants were sensitive to the quality of the advice (see also Yaniv & Kleinberger, 2000). These results square with the cognitive account, whereby people's weighting of the advice is a function of the strength of the evidence retrieved for their own opinion and for the advice.

Whereas the cognitive account is sufficient to explain the results, self-serving motivations (e.g., prior commitment) may further contribute to egocentric discounting. The need for consistency might be present even under conditions that reduce commitment pressures, as in the present studies. It is not easy however to delineate its exact role, as the following study suggests. In one condition, Sniezek and Buckley's (1995) respondents formed their prior opinions on binary-choice questions *before* getting the advice (answers) of two advisors. Other respondents in another condition formed their opinions on the same questions while being presented with the others' advice. Respondents were slightly less likely to change their opinion if they had formed an independent opinion than if they had not. While this finding could easily be explained in terms of commitment to prior opinion, an alternative cognitive account could also be offered. Respondents who had formed an independent opinion retrieved from memory supporting information that was richer and more elaborate than the information that they would have retrieved had the advisory opinions been presented to them concurrently.

Perhaps rather than trying to distinguish between the cognitive and the motivational accounts, it would be useful to investigate how *adding* commitment pressures (e.g., ego threats, the need to save face) affects advice discounting. A plausible hypothesis is that such pressures would decrease the utilization of advice. Further research should shed light on this hypothesis and on how commitment and consistency affect the gains from using advice.

Implications

Regardless of the position one takes on the roles of motivation and cognition, the present findings have important prescriptive implications to offer. Clearly people extract useful information from advice—their final self-estimates are more accurate than the initial ones.

The presumed advantage of advice—as a source of potentially disconfirming information—has been proven correct. Schotter (2003) also observes that "most of the time we make decisions relying only on the rather uninformed word-of-mouth advice we get from our friends or neighbors." He calls this naïve advice. Schotter's game-theoretical experiments, which involved naïve rather than expert advice, also suggest that participants behave more rationally when they make decisions under the influence of advice (Schotter, 2003). It is important to reiterate that to be helpful advice does not need to arrive from more knowledgeable individuals—just (fully or partially) independent ones!

Notwithstanding the benefit of using naïve advice, people do not extract all the information contained in the advice. Most of the simple formal combination rules that we have examined (including equal weight, median, and consensus trim) outperformed the human decision-makers, suggesting that *consistent* application of rules is superior to intuitive revision (which tends to be inconsistent), in agreement with earlier research on linear models of prediction (cf. Dawes, 1979; Einhorn, 1972). A final prescriptive implication is that the judicious removal of extreme opinions (using consensus-based trimming or the median) could boost the accuracy of global judgments.

Aside from the prescriptions for better utilization of advice, these results also have a broader theoretical message. A fundamental issue in the literature in recent decades has been the rationality of human behavior, in light of the research showing flaws in intuitive judgment and decision making. A pervasive impediment to human judgment appears to be the difficulty of generating alternatives to one's current thoughts. Framing effects, anchoring, and confirmation bias all arise when people fail to generate relevant alternatives. Seeking and utilizing advice can be viewed as a corrective procedure, or an adaptive *social decision-support system* that helps individuals encounter new alternatives. Advice, be it a new anchor, a different frame, or a piece of disconfirming information, can trigger beneficial thought processes.

Reservations and limitations

Our study investigated only a small set of possible revision rules out of a large number of possibilities. The rules we evaluated here are static and global rather than dynamic and contingent; they were applied universally, regardless of context, and had no "memory" for past successes or failures. Our results are therefore only an approximation of the underlying processing of opinions. Might dynamically complex rules with contingencies lead to deeper insights? We cannot rule out this possibility, though it is also plausible that complex rules would be more costly in terms of added parameters, as well as difficult to interpret. Research on judgment analysis has pointed out that simple models of judgment (of the sort used here) are rarely surpassed by more complex ones (Einhorn & Hogarth, 1975).

Our final comment concerns the limited scope of the advice studied here. We only considered *quantitative* advice. Whereas numerical forecasts are often important in real life, everyday decision making also involves *qualitative* advice. Moreover, advice is sometimes presented along with supporting arguments. Future research should focus on the integration of other advice formats, such as probabilities (Yates, Price, Lee, & Ramirez, 1996), preferences, and advice supported by arguments (Jonas & Frey, 2003). While we anticipate that the basic revision mechanisms found here will also be relevant to other sorts of advice, we believe that new research will shed more light on advice-based decision processes.

References

Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners*. Dordrecht, Netherlands: Kluwer.

Aronson, E., Turner, J., & Carlsmith, M. (1963). Communicator credibility and communicator discrepancy as determinants of opinion change. *Journal of Abnormal and Social Psychology, 67*, 31–36.

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

Bochner, S., & Insko, C. A. (1966). Communicator discrepancy, source credibility, and opinion change. *Journal of Personality and Social Psychology, 4*, 614–621.

Brockner, J., Rubin, J. Z., & Lang, E. (1981). Face-saving and entrapment. *Journal of Experimental Social Psychology, 17*, 68–79.

Brown, R. (2000). *Group processes*. Malden, MA: Blackwell.

Brunswik, E. (1955). Representative design and probabilistic theory. *Psychological Review, 62*, 193–217.

Budescu, D. V., Rantilla, A. K., Yu, H.-T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes, 90*, 178–194.

Camerer, C. F., & Johnson, E. J. (1991). The process–performance paradox in expert judgment: How can experts know so much and predict so badly? In A. Ericsson & J. Smith (Eds.), *Toward a general theory of Expertise: Prospects and limits* (pp. 195–217). New York: Cambridge University Press.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*, 813–838.

Cialdini, R. B. (1993). *Influence: The psychology of persuasion*. New York: William Morrow.

Clemen, R. T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting, 5*, 559–583.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego: Academic Press.

Davis, J. H., Zarnoth, P., Hulbert, L., Chen, X.-p., Parks, C., & Nam, K. (1997). The committee charge, framing interpersonal agreement, and consensus models of group quantitative judgment. *Organizational Behavior and Human Decision Processes, 72*, 137–157.

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

DeGroot, M. H. (1986). *Probability and statistics* (second ed.). Reading, MA: Addison-Wesley.

Dunning, D., & Hayes, A. F. (1996). Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology, 71*, 213–229.

Einhorn, H. J. (1972). Expert measurement and mechanical combination. *Organizational Behavior and Human Performance, 7*, 86–106.

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.

Gardner, P. H., & Berry, D. C. (1995). The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology, 9*, s55–s79.

Harries, C., Yaniv, I., & Harvey, N. (2004). Combining advice: the weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making, 17*, 333–348.

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

Harvey, N., Harries, C., & Fischer, I. (2000). Using advice and assessing its quality. *Organizational Behavior and Human Decision Processes, 81*, 252–273.

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

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: the belief-adjustment model. *Cognitive Psychology, 24*, 1–55.

Jayaraman, K. S. (2005). Rival monsoon forecasts banned. *Nature, 436*, 161. doi:10.1038/436161.

Johnson, T. R., Budescu, D. V., & Wallsten, T. S. (2001). Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value. *Journal of Behavioral Decision Making, 14*, 123–140.

Jonas, E., & Frey, D. (2003). Information search and presentation in advisor–client interactions. *Organizational Behavior and Human Decision Processes, 91*, 154–168.

Kennedy, J., Kleinmuntz, D. N., & Peecher, M. E. (1997). Determinants of the justifiability of performance in ill-structured audit tasks. *Journal of Accounting Research, 35*, 105–123.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 461–469.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221–232.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences—effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19*(5), 448–468.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

Kunda, Z., & Oleson, K. C. (1997). When exceptions prove the rule: how extremity of deviance determines the impact of deviant examples on stereotypes. *Journal of Personality and Social Psychology, 72*, 965–979.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: misappreciation of the averaging principle. *Management Science, 52*, 111–127.

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.

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.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156–166.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behavior and Human Performance, 167*, 366–387.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inferences: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297–340). New York: Academic Press.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40*, 193–218.

Schotter, A. (2003). Decision making with naive advice. *American Economic Review, 93*, 196–201.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven, CT: Yale University Press.

Slovic, P. (1966). Cue-consistency and cue-utilization in judgment. *The American Journal of Psychology, 79*, 427–434.

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.

Soll, J. B. (1999). Intuitive theories of information: beliefs about the value of redundancy. *Cognitive Psychology, 38*, 317–346.

Sorkin, R. D., Hayes, C. J., & West, R. (2001). Signal detection analysis of group decision making. *Psychological Review, 108*, 183–203.

Streiner, D. L. (2000). Do you see what I mean? Indices of central tendency. *Canadian Journal of Psychiatry, 45*, 833–836.

Sunstein, C. R. (2003). *Why societies need dissent*. Cambridge, MA: Harvard University Press.

Surowiecki, J. (2004). *The wisdom of crowds*. Random House.

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.

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology, 37*, 62–69.

Wilcox, R. R. (1992). Why can methods for comparing means have relatively low power, and what can you do to correct the problem? *Current Directions in Psychological Science, 1*, 101–105.

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

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83*, 260–281.

Yaniv, I. (2004a). The benefit of additional opinions. *Current Directions in Psychological Science, 13*, 75–78.

Yaniv, I. (2004b). Receiving other people's advice: influence and benefit. *Organizational Behavior and Human Decision Processes, 93*, 1–13.

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.

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