Journal of Behavioral Decision Making
J. Behav. Dec. Making, **17**: 95–116 (2004)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.464

Weighting Information from Outside Sources: A Biased Process

SHENGHUA LUAN, ROBERT D. SORKIN* and JESSE ITZKOWITZ
University of Florida, Gainesville, Florida, USA

ABSTRACT

How do people utilize information from outside sources in their decisions? Participants observed a signal-plus-noise or noise-alone event and then made a yes–no decision about whether a signal had occurred. Participants were provided with two information sources to aid decision making. Each source consisted of four components that provided estimates of signal likelihood. In Experiment 1, the two sources had equal overall accuracy but differed in the expertise and internal correlation of their components. A regression analysis indicated that participants overweighed the high-expertise-high-correlation source. This bias occurred on trials when the aggregate opinions of the sources disagreed. In Experiment 2, both the overall accuracy of the source and its components were manipulated. Participants overweighed information from the higher accuracy source. These biases reflect people's sensitivity to across-trial and within-trial differences in the accuracy and internal consistency of information sources. Experiment 3 provided additional evidence supporting these conclusions. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS decision weights; decision bias; information sources; belief updating; evidence weighting

People often make decisions using information and advice from other sources. For example, imagine that you are CEO of a firm that manufactures several lines of children's clothing. Your firm has been considering introducing a completely new product line. In addition to the obvious risks, such a move would involve a sizable commitment in advertising and other expenses for your company. You have asked two marketing consulting firms to assist you in that decision. Firms A and B each study the question and make separate presentations. The first presentation is made by four experts from Firm A. These experts arrive separately at your office and provide a varied set of views and predictions. Their recommendation, while not unanimous, is to postpone the decision for a year. The next day, Firm B's experts make their recommendation. They arrive together and their presentations seem more rehearsed than A's and are consistent with each other. It is clear that the Firm B experts conferred before arriving at your office. Firm B's recommendation is to go

*Correspondence to: Robert D. Sorkin, AFOSR/NL Perception & Cognition, 4015 Wilson Avenue., Rm. 713, Arlington, VA 22203-1954, USA. E-mail: robert.sorkin@afosr.af.mil

Contract/grant sponsor: Air Force Office of Scientific Research; contract/grant number: F496200000100041.

ahead with introducing the new product line this year. Which firm's recommendation weighs more heavily in your decision? And how did you form that weighting strategy?

To answer these questions, we asked people to perform a signal detection task while aided by information sources that were composed of multiple components. Of main interest was how participants weighed the information received from different sources as a function of the information value (or aggregate accuracy) of the sources. We also were interested in the effects of the statistical properties of the source's components. We wondered, if all other factors were set equal, whether people would exhibit a bias toward the more internally consistent sources.

One might expect that the statistical properties of an information source should influence the weight a decision maker assigns to estimates received from that source. One obvious property is the information value of the source. From a normative perspective, sources that are more capable of providing accurate information about a decision should be considered more important, because they will more often lead to a correct decision. Assuming that a person is a rational decision maker, the decision weight assigned to an information source should be proportional to the expertise or actual competence of the source (see, Grofman, Feld, & Owen, 1984; Nitzan & Paroush, 1982; Shapley & Grofman, 1984; Sorkin & Dai, 1994; Sorkin, Hays, & West, 2001; Wallsten, Budescu, Erev, & Diederich, 1997). There have been many studies about the effect of ''source-credibility'' (Birnbaum & Stegner, 1979; Chaiken & Maheswaran, 1994; Fishbein & Ajzen, 1975). These studies have generally found that advice from sources consisting of members with high credibility (i.e., greater expertise) was more influential on people than advice from sources composed of members with lower credibility (less expertise).

Another obvious property of sources is the number of distinct information components that they contain. Suppose that two sources A and B are composed, respectively, of five and eight components. Assume that all of A's and all of B's components are statistically independent and that each of A's sources are equal in competence to each of B's components. Normatively, we should assign a greater weight to source B, because information from B's aggregate opinion is derived from three more components than A. On the other hand, suppose that the two sources have equal aggregate expertise. Since A has fewer components, each of A's components must have somewhat greater competence than each of B's components. Because the information values are equal, there should be no difference in the weight given to an opinion from A or B.

A study conducted by Harkins and Petty (1981) addressed these issues. They found that information presented by multiple sources received greater scrutiny than the same information presented by a single source. In other words, information from multiple sources enhances message processing, and the arguments the sources present may have greater potential to influence people. In a subsequent study (1987), Harkins and Petty explored why multiple sources had this power. In one experiment, they found that the persuasive advantage of multiple sources was eliminated when the members that composed these sources were said to have formed a committee rather than being independent. In another experiment, when this committee was believed to include members with dissimilar perspectives, the advantage of multiple sources was retained. Based on these findings, they concluded that the power of multiple sources resides in perceived informational independence and the divergent perspectives such sources are presumed to represent.

Two problems with these studies are the lack of parametric manipulation of the properties of the different sources and the lack of measurement of the resultant decision weights. In these studies, only one information-relevant property of a source was manipulated. In order to hold the information value of a source constant, one would have to vary at least two properties of the source (i.e., increase the number of components of a source while simultaneously decreasing the expertise of each component). A third problem is that differences in the source properties were either very conspicuous (like the number of members in a source in Harkins and Petty, 1987), or explicitly told to the participants before the task (like the source credibility information in Birnbaum and Stegner, 1979). We wondered whether (paid and appropriately trained) participants would be sensitive to the appropriate statistical properties of the information sources, rather than simply respond to explicitly cued differences in the sources as in the previous studies.

An analysis by Sorkin and Dai (1994) provides a quantitative method for specifying the information value of an information source based on the statistical properties of the source components. Using the theory of signal detection (Green & Swets, 1966), Sorkin and Dai analyzed the behavior of a theoretically optimum group of decision makers in a signal detection task. Their model specifies the highest achievable detection performance $d'_{ideal\text{-}group}$ that is obtainable from an array of m group members, having specified individual indices of detection expertise $\{d'_i\}$ and a uniform pair-wise correlation ρ among the members' judgments:

$$d'_{ideal\text{-}group} = \left[\frac{m\sigma^2_{d'}}{1-\rho} + \frac{m(\mu_{d'})^2}{1+\rho(m-1)}\right]^{1/2} \tag{1}$$

where $\sigma^2_{d'}$ is the variance of the array of members' expertise and $\mu_{d'}$ is the average detection ability of those members. It can be seen that ideal group performance is a decreasing function of the correlation and an increasing function of the group size and member expertise, which is consistent with other group or expert decision-making models (Ashton, 1986; Hogarth, 1978; Clemen & Winkler, 1999).

Rather than considering the performance of a *group*, in this paper we use Equation 1 to specify the performance of an information *source* consisting of *m components* having properties $m,\ \rho,\ \sigma^2_{d'}$, and $\mu_{d'}$. That is, we equate an information source that contains m components to a group that contains m members. Second, we assume that the value of information from a source is indicated by how often that source makes correct decisions. The appropriate measure of this performance is given by the source's ideal detection index as specified by Equation 1: the higher the value, the more likely that the source's information will result in accurate decisions.

Consider a decision maker who is faced with a decision problem and has received advice from two different and independent sources: $A_{estimate}$ and $B_{estimate}$ such as from Firms A and B in the beginning example. This decision maker should form a decision statistic Y that is the linear sum of the weighted estimates from the two sources, where each source's estimate is weighted by that source's aggregate detection index (Sorkin & Dai, 1994):

$$Y = d'_{source\text{-}A} A_{estimate} + d'_{source\text{-}B} B_{estimate} \tag{2}$$

If the detection index is equal for two such sources, the information received from them has equal value, and the decision maker should weigh that information equally. Any deviation from these ideal weights will result in a less accurate decision statistic and poorer performance.

Suppose that Source A has an information value $d'_A = 2.0$ and Source B has an information value $d'_B = 1.5$; normatively, their relative weights in a decision maker's final decision should be 4 : 3. If a different weight ratio is employed, the decision maker's maximum accuracy will be decreased (Sorkin & Dai, 1994; Sorkin *et al.*, 2001). After assessing the decision weights that a decision maker actually assigns to differing sources by using regression or correlation techniques (see Methods), we can determine any deviation from optimal behavior. Thus, in addition to allowing parametric variation of the statistical properties of information sources and thus their aggregate information value, Equation 1 makes it possible to draw quantitative inferences about whether and to what extent a participant deviates from an optimal weighting of an information source.

Because of its proved usefulness in multiple-cue learning and utilization research, which shares the similar task structure to the above source utilization problem, a lens model analysis may also be employed to draw inferences about the participant's weighting strategy (Cooksey, 1996; Lee & Yates, 1992; Tucker, 1964). To apply the lens model analysis, one would consider Source A and B's estimates as information "cues" on which the judgment or decision is based, the cue's ideal weights as the cue validity coefficients, and their obtained weights as the cue utilization coefficients. Although the lens model approach has been found most useful for assessing the achievement or agreement of human judgment systems (Brehmer & Joyce, 1988), we thought that a signal detection approach would offer more insight into how the inherent statistical properties of the sources would affect the relative weighting of the information sources.

When information from multiple sources is offered to a decision maker, should we expect to observe optimal weightings by the decision maker? The ability to weigh information in an optimal fashion may be mitigated by a number of factors. First, a decision maker may not know how to evaluate the information value of a source or may lack sufficient familiarity with that source. Indeed, it is often difficult to ascertain who the expert is in a given situation. Additionally, deviations from ideal behavior may occur even though a decision maker is able to determine the information values of the sources because s/he may simply lack the resources or motivation to perform these calculations. According to the ‘‘cognitive miser’’ perspective, people spend less cognitive resources when ‘‘heuristics’’ or ‘‘shortcuts’’ are available to help them to solve the problem (Taylor & Fiske, 1978). This tendency is so strong that it occurs even when people have enough time and resources to think about the problem carefully. This behavior also has been described well by the ‘‘satisficing strategy,’’ proposed by Simon (1956). If people can obtain solutions that make them satisfied using less effort, they will not expend more effort finding ‘‘optimal’’ ones. A similar view is expressed by the ‘‘fast and frugal’’ models proposed by Gigerenzer and his colleagues (1999). People sometimes make their decisions based on only one cue, rather than searching the whole information space and integrating all of the relevant information present. Decisions based on these fast and frugal heuristics can often be as good as applying more complicated methods. Thus, if simple strategies exist that help a decision maker evaluate the importance of a source's advice, it is very possible that s/he will use information recruited by these simpler strategies, which may or may not accurately reflect the sources' actual information values.

In order to directly study these questions, we placed participants in a traditional signal detection experiment in which they were asked to judge whether a visual display indicated that a noise event or a signal-plus-noise event had occurred. To aid the participants in their detection task, we provided two additional information sources, Source A and Source B, each composed of four components. After the participant had observed the input on each trial, Sources A and B each provided the participant with four continuous ratings that estimated the occurrence of a signal-plus-noise event on that trial. Thus, in addition to the participant's own estimate of signal-plus-noise, $P_{estimate}$, the participant now had estimates $A_{estimate}$ and $B_{estimate}$ to form a final decision statistic. Assuming that these three estimates were independent and were weighed by the participant in an optimal fashion, the final index of the participant's detection performance in the task would be:

$$d'_{participant} = [(d'_P)^2 + (d'_A)^2 + (d'_B)^2]^{1/2} \tag{3}$$

where d'_P is the index of detection of the participant *without* the aid of Sources A and B, and d'_A and d'_B are the detection indexes of each source alone.

In Experiment 1, we manipulated the statistical properties of these sources so that the information value of the two sources was equal, $d'_A = d'_B$, but the sources had different average component detection indices $\mu_{d'}$ and different component correlations, ρ. That is, using Equation 1 we created pairs of sources that had equal information value but different levels of correlation and component expertise; whereas one source had a higher mean expertise μ'_d and higher internal correlation ρ, the other had a lower μ'_d and lower ρ. One source had components that were *individually* ‘‘better’’ at the task, but these estimates were partially correlated with the estimates of other components within the source. Thus, the overall performance of this source was equal to the other source whose components had lower individual d' but were independent in their estimates. Participants were not told the differences between the two sources. We assessed the decision weights employed by the participants by calculating a multiple regression of each participant's final decisions on the participant's initial estimates and the estimates generated by each source.

EXPERIMENT 1: EQUAL INFORMATION SOURCES

Method

Participants were presented with a multiple-element visual display consisting of nine analog gauges, similar to those shown in Figure 1. The values displayed on the nine gauges were determined by sampling from one of two normal distributions: for signal-plus-noise, $\mu_s = 5$, and for noise-alone, $\mu_n = 4$. Both distributions had a common standard deviation, $\sigma_{com} = 2.5$. After viewing the display, participants made an estimate, via a response slider, of the likelihood that the display of the nine gauges indicated the occurrence of a signal-plus-noise or noise-alone event on the trial. Following this estimate, two simulated information sources were displayed as shown in Figure 2. Each source consisted of a group of four information components, each providing estimates of the likelihood of signal occurrence on that trial. After presentation of the information sources, the participant was required to make a yes–no decision about the occurrence of signal on that trial. A monetary payoff was contingent on the accuracy of this yes–no decision. Based on previous experiments using similar graphical materials (Montgomery & Sorkin, 1996), the expected performance d' of our participants in the basic detection task (i.e., without the two information sources) was approximately 1.0.

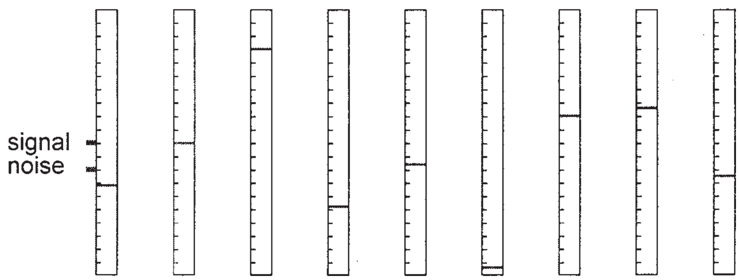

Figure 1. This shows an example of the stimulus array presented to a participant on a noise-alone trial of the experiment. On each trial, the values on the nine gauges were drawn from either the signal-plus-noise or the noise-alone distributions. The thick ticks labeled "signal" and "noise" indicate, respectively, the means of these distributions. The value of the common standard deviation was 2.5.

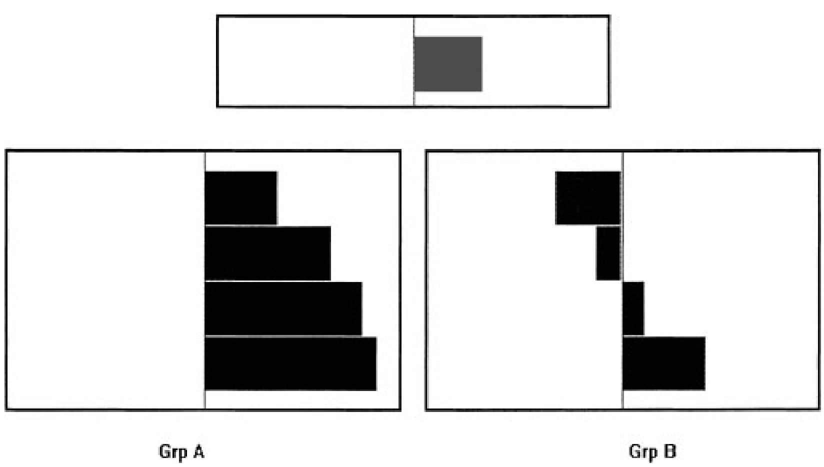

Figure 2. This shows an example of the display of the source information, with the participant's initial judgment on top (displayed to participant in green). The left source was always marked as Source A (displayed in blue) and the right source was always marked as Source B (displayed in red). In each source, component opinions were ordered by their estimates of the likelihood of signal-plus-noise on that trial (on a 0 to 100 scale).

Participants

Five University of Florida students, three women and two men, participated in the study. All participants had normal or corrected-to-normal visual acuity. Participants were paid \$5.50 per hour plus an incentive bonus that was based on the accuracy of performance, which ranged from \$1 to \$2 per hour.

Apparatus and stimuli

Stimuli were generated, presented, and responses recorded using Pentium® computers. The stimuli were displayed on a 15-inch color monitor (60 Hz refresh rate at 1024×768 resolution). Subjects sat approximately 60 cm away from the monitor. The nine gauges subtended a visual angle of approximately 8° vertical by 22° horizontal. Responses were made via a standard computer keyboard. Individual display elements, shown in Figure 1, consisted of two parallel vertical lines with tick marks on the left line, dividing the gauge into 20 intervals (from 0 to 10). Two larger tick marks on the leftmost gauge marked the means of the noise and signal-plus-noise distributions. On a given trial, all of the elements displayed values that had been drawn independently from the same distribution. Half of the trials (randomly) were from the signal-plus-noise and half from the noise-alone distribution. Stimulus duration was 360 ms during all experimental trials. A centered cross (0.5 inch) fixation stimulus preceded the stimulus (200 ms), and a white masking screen (200 ms) followed presentation of the stimulus.

After presentation of the stimulus, the participant made an initial judgment about signal occurrence. This judgment was made on a mouse-operated slider and consisted of a rating, ranging from 0 to 100 from left to right, of the estimated likelihood of signal occurrence on that trial. The line in the middle marked 50, which was the neutral point. After making this estimate, estimates from two simulated information sources (shown in Figure 2) were displayed together with the participant's initial judgment. The source information consisted of ratings similar in format to the participant's initial judgment (i.e., ranging from 0 to 100 from left to right) of the estimated likelihood of signal occurrence on that trial. The upper box contained the participant's rating and the two boxes below contained the ratings from the two sources. Each source was composed of four simulated components and each component within the same source had the same detection ability ($\sigma^2_{d'} = 0$ and $\mu_{d'} = d'_i$). We shall refer to the average of the component detection indices of each source, respectively, as d'_a and d'_b, in order to distinguish those indices from the overall (aggregate) source detection indices d'_A and d'_B.

The simulated ratings for the components of an information source were generated from two normal distributions. On signal-plus-noise trials, the average rating was higher than 50 and for noise-alone, it was lower than 50. On both types of trials, these generated ratings had a common standard deviation of approximately 16. The difference between the means of the two distributions was manipulated to control the component's detection index. A component rating from a given source was independent of the ratings from the other source and from the participant's displayed input. However, the component ratings within a source were partially correlated with each other, depending on the experimental condition. This internal correlation was manipulated by a method described by Sorkin (1990): A component's rating (within a source) was generated by:

$$R_i = aX_i + bX_c \tag{4a}$$

and

$$R_j = aX_j + bX_c \tag{4b}$$

where a and b are constants and X_i, X_j, X_c are independent, zero mean, equal variance, normal random variables. The correlation between any pair of components of that source is given by:

$$\rho_{R_i,R_j} = \frac{b^2}{a^2 + b^2} \tag{5}$$

Within a source, the pair-wise correlations between components were equal. Thus, within any trial, the component estimates from the higher correlation source were more consistent with each other than were those from the lower (or zero) correlation source.

Procedure

Following presentation of the stimulus, the participant was required to give a rating ranging from 0 to 100 indicating his/her estimation of the likelihood of a signal-plus-noise occurrence in that trial. No reward or penalty about the correctness of this initial judgment was applied to this estimate. After the estimate was made, the simulated sources' ratings plus the participant's rating were displayed and the participant was required to make a final yes–no response. If the final decision were correct, a 2-cent bonus was given; otherwise, a 2-cent penalty was applied. Full feedback about the correctness of the final decision was supplied after the decision. Participants could take as much time as they wanted to make the initial estimate or the final decision. Each participant performed 16 blocks of such trials consisting of 100 trials in each block.

There were two experimental conditions with eight blocks of trials per condition. In Condition I, Source A had higher component detection abilities ($d'_a = 0.96$) as well as higher component correlations ($\rho = 0.2$) than Source B ($d'_b = 0.75$ and $\rho = 0$). Due to the relatively small difference between the two sources' detection abilities and component correlations, we will refer to Condition I as the "small difference" condition. Within this condition, both sources had the same overall information value or overall detection performance (i.e., $d'_A = d'_B = 1.5$). Condition II, the "large difference" condition, had the same design as the previous condition. However, Source A had even higher component abilities ($d'_a = 1.11$) and higher component correlations ($\rho = 0.4$) than in the previous condition. Again, the overall information value of the two sources was equal. In order to counterbalance any position effect, Source A was displayed on the left for four blocks in each condition and Source B was displayed on the left for the other four blocks. In all the displays, the label "Grp A" was always in the left (see Figure 2). The order of the sixteen blocks was randomized for each participant. Table 1 gives a summary of these experimental conditions.

Participants received intensive training (at least 1000 trials) before the experimental sessions. During these training trials, no information sources were present to assist the final decision. Instead, after rating the likelihood of a signal-plus-noise event, they were immediately asked to make a "yes–no" decision. Participants were encouraged to rest between blocks and could voluntarily rest during any trial by intentionally delaying their response. When aided by the information sources, it took a participant about 15 seconds to finish an experimental trial and 25 minutes to complete a block of 100 trials.

Prior to running the experimental sessions, participants were told that the accuracy of the source information depended on two properties of the source: its component detection abilities and the internal correlation

Table 1. Source conditions in Experiment 1: equal information sources

	Source A			Source B		
Condition	Component expertise d'_a	Component correlation ρ	Source expertise d'_A	Component expertise d'_b	Component correlation ρ	Source expertise d'_B
I: Small difference	0.96	0.2	1.50	0.75	0	1.50
II: Large difference	1.11	0.4	1.50	0.75	0	1.50

of its components' estimates.[1] They were told that the two sources might be different in one or more of these properties and were asked to pay attention to these properties. By this instruction, we hoped to have participants focus on the two statistical aspects of the sources, thus minimizing the influence of any irrelevant source properties.

Results

We first verified that the statistical properties (component d' and ρ values) of each source were close to the quantities specified by our manipulation. This proved to be the case,[2] indicating that the manipulations were successful and that the two simulated sources had an equal information value (or equal source expertise). As an additional check, we evaluated the aggregate detection performance of the sources (A and B) by averaging the four ratings from each source's components and then partitioning the average rating response of each source into a (simulated) "yes" or "no" response. (A rating greater than 50 was deemed a "yes" response.) The resulting hits and false alarms were then tabulated and an aggregate d' value was computed for each source (Macmillan & Creelman, 1991). In both conditions and for each participant, the aggregate d' of each source was very close to the target value of 1.5 and the ratios of these d's were very close to 1. We also evaluated the participants' unaided detection accuracy by converting their initial ratings to simulated yes–no responses; a rating of 50 or more was deemed a "yes" response. In the small and large difference conditions, respectively, the average detection accuracy of the participants was 0.90 ($sd = 0.09$) and 0.92 ($sd = 0.16$), consistent with prior experiments with these types of displays (Montgomery & Sorkin, 1996).

Analysis of decision weights

In order to assess the decision weights that participants assigned to each source, we constructed a binary logistic regression model of the participant's final yes–no decision.[3] In the regression model, the dependent (predicted) variable was the final decision (equal to 1 for a "yes" or 0 for a "no"), and the independent variables (predictors) were, respectively, the participant's initial estimate, the average estimate of Source A's four components, and the average estimate of Source B's four components. Standardized regression coefficients, the beta weights (β_i), were obtained to measure the weights assigned to each predictor.

All β-weight values resulting from each participant's regression model were significantly greater than zero ($p < 0.001$), indicating that estimates from either Source A or Source B were significant in forming each participant's final decision. A Hosmer and Lemeshow goodness-of-fit test was performed to examine how

[1]Instructions about source properties:

"The two simulated sources you will see may be different from each other in two properties: the detection abilities of its components and the information correlation among its components. "Component detection ability" represents how capable every component of the source is to make a correct judgment. Inside a certain source, all of its components have the same detection abilities. But in each block, Source A and Source B may have the same or different component abilities. The second property is the "information correlation among the components." This property represents how much the information every component of the source uses to make their estimates is correlated with each other. The more they are correlated, the more similar their estimates are. Again, Source A and Source B may be the same or different in this aspect. I know you may not fully understand the second property. Whatever your understanding is, keep it in mind when you try to take these sources' estimates as references to make your final decisions, because it is a factor that can contribute to how accurate the judgment of the simulated source is, just like the property of component detection ability."

[2]The values obtained in all the conditions of Experiments 1 and 2 were very close to the theoretical values, with very small variance. These values, as well as the statistical summaries and raw data, are available on request from the authors.

[3]In logistic regression models, the predicted variable is in the form of natural logarithm of an odds ratio:

$$\mathrm{Log}\left(\frac{\pi(x)}{1-\pi(x)}\right),$$

in which $\pi(x)$ is the probability of a signal-plus-noise occurrence and $1 - \pi(x)$ is the probability of a noise-alone occurrence. A correlation technique developed by Richards and Zhu (1994) to assess decisional weights in SDT experiments was also applied and the results produced using their method were essentially identical to those obtained by logistic regression analysis.

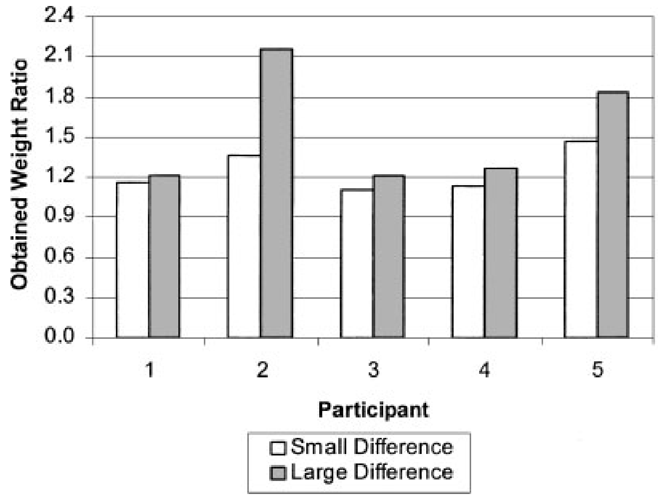

Figure 3. The obtained weight ratios between Sources A and B (β_A/β_B) from the multiple regression analysis for each participant in the two experimental conditions in Experiment 1: equal information sources.

well the obtained regression model predicted the dependent variable for each participant in the two experimental conditions. This test was non-significant ($p > 0.05$), indicating that the models could provide good predictions of the final decisions made.

To measure the relative impact of the two sources, we calculated the ratios between the obtained weights assigned to them (β_A/β_B). These ratios were all larger than 1, indicating that all participants assigned greater weight to Source A than to Source B when forming their final decisions. The obtained weight ratios in the two experimental conditions for each participant are illustrated in Figure 3. Note that although there were individual differences, all the participants had the same pattern of weighting in the two conditions.

Readers who are unfamiliar with psychophysical methodology may find it unusual to encounter a decision experiment that employs only five participants. The difference between our experiment and typical judgment experiments that employ five times as many participants is in the number of trials per participant—in our case more than 2600 trials per participant or a total of approximately 13 000 trials (including training). The relatively large number of trials per participant is necessitated by use of the signal detection paradigm. Basically, one trades some loss of generality across participants for greater measurement precision. The gamble in this approach is that one will obtain a consistent *pattern* of behavior among the participants. Since that is what we found—a consistent pattern of source weighting for each of the five participants—we would argue that the present approach is as valid as others that use many times more (possibly unpaid) participants and a fraction of the trials.

Figure 4 allows comparison of the average obtained weight ratios and the corresponding average ideal weight ratios (d'_A/d'_B) for each condition. Statistical tests confirmed that the differences between the observed and ideal ratios were significant ($t(4) = 2.330$, $p < 0.05$, and $t(4) = 2.625$, $p < 0.05$, in the small and large difference conditions respectively). Furthermore, the actual weight ratios obtained in the large difference condition were in all cases larger than those obtained in the small difference condition, indicating that the bias towards Source A increased when Source A had higher component detection abilities and higher internal correlations, although the difference was only marginally significant ($t(4) = 2.058$, $p = 0.054$).

Agreement vs. disagreement

Previous research has found that the weighting patterns were different when sources' or cues' estimates were incongruent or less consistent with each other, compared to when they were congruent or more consistent (e.g. Slovic, 1966; Sniezek & Buckley, 1995). To examine whether the degree of decision conflict between

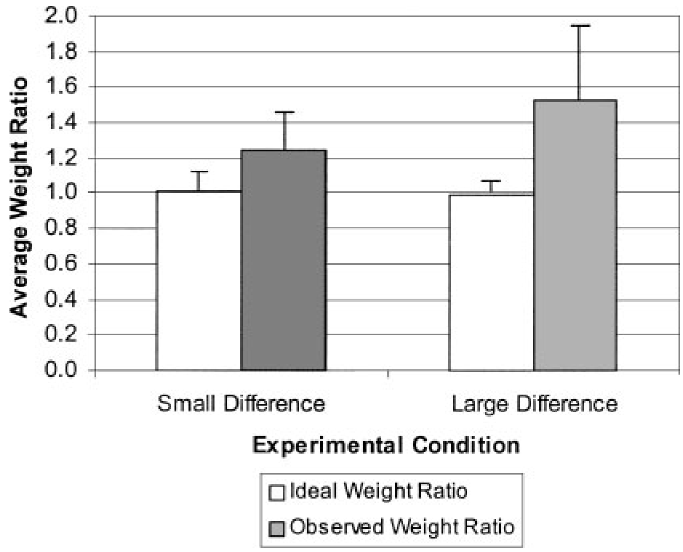

Figure 4. The average ideal and obtained weight ratios between Sources A and B in both conditions of Experiment 1. The ideal weight ratio is the ratio of Source A's information value over Source B's information value (d'_A/d'_B), and the obtained weight ratio is β_A/β_B in the multiple regression analysis. The brackets indicate one standard error of the mean.

sources had an impact on their obtained weight ratios, we partitioned all experimental trials into one of the following two categories: *agreement* trials when Source A's aggregate opinion (average rating > 50 or average rating < 50) agreed with Source B's aggregate opinion; and *disagreement* trials when the two sources disagreed with each other. After sorting the trials in this way (approximately one third of the trials were placed in the disagreement category), the decision weights and weight ratios were then recalculated.

On trials where the two sources agreed with each other, the obtained ratios were not significantly different from 1 ($t(4) = 0.110$, $p > 0.05$ in Condition I and $t(4) = 0.167$, $p > 0.05$ in Condition II). However, on trials where the two sources disagreed, the ratios were larger than 1 ($t(4) = 5.178$, $p < 0.01$ and $t(4) = 4.273$, $p < 0.01$ in the small and large difference conditions, respectively). Additionally, these ratios were higher in the large difference condition than in the small difference condition ($t(4) = 2.528$, $p < 0.05$). The average weight ratios for both conditions are illustrated in Figure 5. These results indicate that the weighting patterns differed according to the degree of conflict between the two sources. Participants' preference for information from Source A was only evident on trials when the aggregate opinions from the two sources were contradictory; and this preference was significantly greater in the large difference condition (Condition II).

Analysis of detection accuracies and efficiencies

Table 2 lists the detection indices (d') obtained from the participants' final decisions, for all experimental trials and for the partitioned agreement and disagreement trials. Regardless of experimental condition, the participants' detection indices were higher on agreement than disagreement trials ($F(1, 16) = 233.19$, $p < 0.001$). Although low, the detection indices obtained on disagreement trials were significantly above the participants' initial detection indices ($F(1, 16) = 8.109$, $p < 0.05$), indicating that some information was being obtained from the information sources.

Detection efficiency, η, is defined by

$$\eta = (d'_{observed}/d'_{ideal})^2 \tag{6}$$

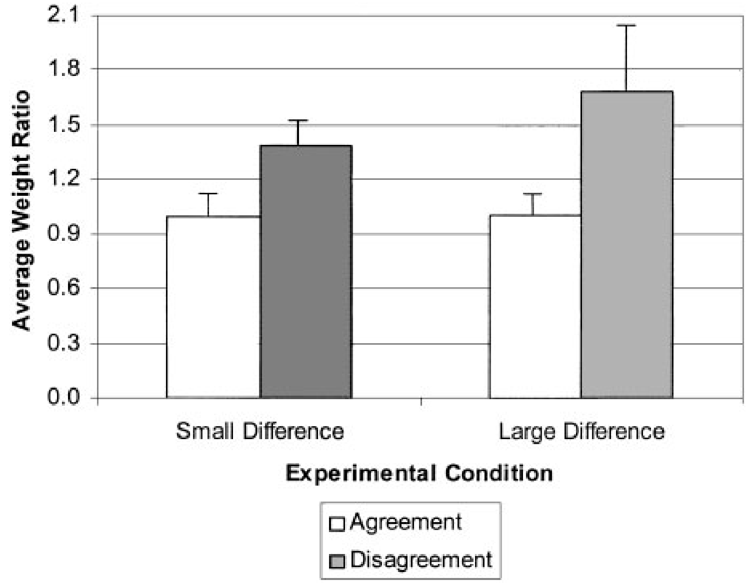

Figure 5. The average obtained weight ratios between Sources A and B when the aggregate opinions of these two sources agreed and disagreed with each other (see text) in Experiment 1. The brackets indicate one standard error of the mean.

Table 2. The detection index d' and efficiency index η of the final decision, calculated on all experimental trials, agreement trials and disagreement trials (see text) in Experiment 1: equal information sources

	Condition I: small difference						Condition II: large difference					
	All trials		Agreement		Disagreement		All trials		Agreement		Disagreement	
Participant	d'	η	d'	η	d'	η	d'	η	d'	η	d'	η
1	2.09	0.69	2.58	0.59	1.44	0.69	2.00	0.71	2.75	0.84	1.09	0.45
2	1.76	0.58	2.42	0.68	0.95	0.38	1.66	0.41	2.04	0.40	1.03	0.35
3	1.97	0.65	2.64	0.73	0.89	0.42	1.88	0.62	2.66	0.77	0.93	0.30
4	2.18	0.81	2.88	0.83	1.37	0.83	1.95	0.67	2.80	0.83	1.00	0.35
5	2.09	0.70	2.68	0.71	1.30	0.52	2.02	0.68	2.64	0.85	1.13	0.46
Mean	2.02	0.68	2.64	0.74	1.19	0.57	1.90	0.62	2.58	0.74	1.04	0.38
sd	0.16	0.90	0.15	0.08	0.23	0.18	0.15	0.12	0.28	0.17	0.07	0.06

Efficiency is a measure of the how much the observed performance differs from the performance of an optimal or ideal detector (Tanner & Birdsall, 1958). In order to calculate the ideal d' values, we summed the ratings of the components of each information source on each trial, and evaluated the detection behavior of each source based on this summed statistic. In order to compute an ideal detection index for the combination of participant and information sources, the resulting detection index was then combined with each participant's initial detection index using Equation 3. This procedure assumes that each decision maker appropriately weighs all the available information from their own observations and the two sources. The efficiency calculation allows a direct comparison of how efficiently information was utilized on these two types of trials, regardless of how much information was available.

The detection efficiencies for each condition are listed within parentheses in Table 2 and the participants' averaged efficiencies are shown in Figure 6. It can be seen that efficiencies were higher on agreement trials

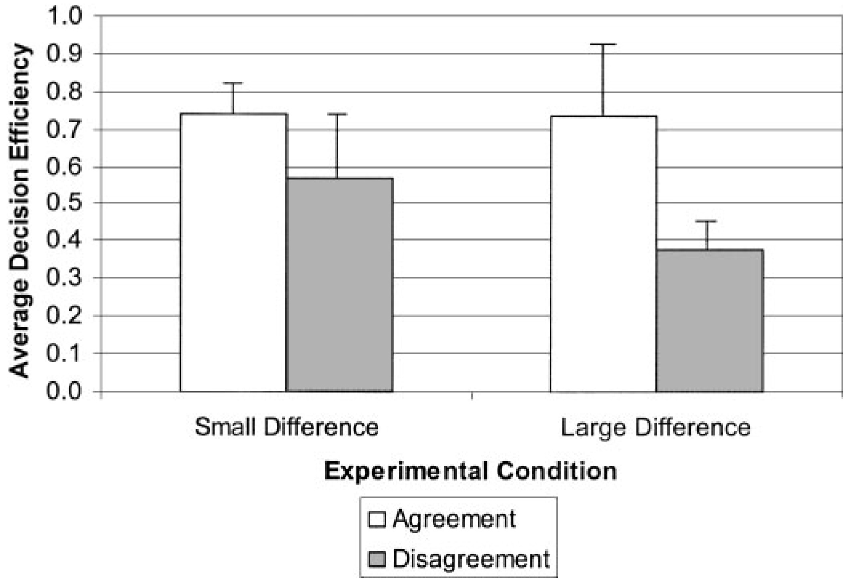

Figure 6. The average detection efficiencies (η) achieved when partitioning trials into agreement or disagreement between Source A and Source B's aggregate opinions in Experiment 1. The brackets indicate one standard error of the mean.

than they were on disagreement trials ($F(1, 16) = 14.18$, $p < 0.01$). There was no difference in the efficiencies between the two experimental conditions on agreement trials ($t(4) = 0.312$, $p = 0.385$), but efficiency was significantly higher in the small difference condition than in the large difference condition on disagreement trials ($t(4) = 2.337$, $p < 0.05$). These results indicate that, on trials when the two sources disagreed, the participants used much less available information from the sources than on agreement trials.[4]

Discussion

It is clear that the decision maker did not assign equal weights to the two sources provided, and that there was a bias towards weighting Source A more heavily in the final decision even though both sources had equivalent information values. Source A, which had higher component expertise and higher component correlation, had a significantly greater influence on the participant's final decision than did Source B, which had a smaller component expertise and smaller component correlation. Moreover, this bias increased when the differences between the properties of the two sources were increased. However, the bias was only exhibited when the average ratings of the components of each source disagreed.[5] In terms of the example at the outset of the chapter, one would have accepted the advice of the experts from Firm B, and decided in favor of offering the new product line.

---

[4]We analyzed part of our data in Experiment 1 using the lens model method, although the analysis was taken in the form of logistic instead of the popular linear regression (Cooksey, 1996), and we had no difficulty producing the typical lens model results, R_a, R_e, R_s, etc. The results were strikingly similar to our SDT analysis. However, they do not seem to provide any additional insight into the core question of why our participants used the information from the two sources (cues) with equal information values (validities) differently in their decision-making process. To avoid being redundant, we chose not to list the results.

[5]An additional reason why we conducted a separate analysis for agreement and disagreement trials was that: in a regression model, when two predictors are highly correlated, it is expected that their resulting weights should be highly similar to each other. Therefore, the ratios of their weights should be close to 1. This was exactly what occurred in the agreement conditions. In disagreement conditions, because the information correlation between Sources A and B was nearly equal to 0, the weights found in a regression model can be unambiguous reflections of their actual weights. As the results indicate, a weighting bias was found in these situations. So, partitioning trials into agreement and disagreement helped us clarify the specific conditions in which participants utilized a biased weighting strategy.

When the aggregate opinions of two sources agree, it is relatively simple and even natural to integrate the information from both sources equally. When the sources disagree, the decision maker is faced with a more difficult situation centering on the question, ''Which source's opinion is better?'' Of course, the decision maker could simply follow her/his own initial estimate and ignore both sources. However, doing so would result in her/his final decision having the same accuracy as her/his initial estimate. Since the participants' final decision accuracy on disagreement trials was higher than the accuracy of their initial decisions, this was not the case. Therefore, the participants must have tried to integrate information from the two sources to help them make a final decision. As shown earlier, the normative weighting strategy assigns equal weight to both sources, even in disagreement cases. Thus, the differential weighting of the sources is a manifestation of a bias that results in lower than optimal performance. When the participant harbors a bias towards one source (giving it more weight than normatively appropriate), s/he does not use all the information provided to her/his, and as a result her/his efficiency is significantly reduced on disagreement trials. The further that the decision maker drifts from an equal weighting strategy, the less efficient her/his performance becomes, as shown in a comparison of the two experimental conditions.

Although a decision bias was found in the experiment, it is difficult to tell which property of the information source–component expertise, component correlation, or both—had the greatest effect on the decision maker's weighting strategy. It is possible that a decision maker could make choices by relying on one ''primary'' source property (Tversky, Slovic, & Sattath, 1988), and this primary property was instrumental in causing the decision maker to weigh the information from the two sources differently. However, because of the simultaneous manipulation of the components' expertise and correlation properties needed to create sources of equal information value, it was impossible for us to specify which property was primary for decision makers in Experiment 1.

It is useful to consider how these two properties might reveal themselves to the decision maker. First, on any trial the component correlation may be detectable because of the greater consistency of the component estimates from within the more highly correlated source. Second, in contrast to the correlation property, differences in the detection accuracy of a component could be deduced over a set of trials using the feedback provided about each trial's outcome. That is, the more expert components would be associated with more correct estimates. Such across-trial information could directly inform the participants of accuracy differences and thereby create a bias to weigh sources that possessed more accurate components more heavily than sources that contained less accurate (but uncorrelated) components. Note that this is in spite of the fact that the accuracy of each source—considered as an aggregate (or average) of its components—was equal.

According to Birnbaum et al.'s configural weighting models (Birnbaum & Stegner, 1979; Birnbaum & Zimmermann, 1998), the weights a decision maker assigns to multiple sources are not only determined by those sources' information values (which is usually thought of as either aggregate expertise or credibility), but also may be affected by the relationships among those sources' estimates, or estimate configurations. In the Birnbaum studies, configurations were primarily constructed on the basis of the scale values of the estimates. The weight that was assigned to each source changed as a function of the relationship between the decision maker's estimate and the estimates of the other sources. Borrowing from their concept, we can see that two ''configural'' differences existed between the sources in Experiment 1: the relative level of estimate consistency and the relative level of component accuracy. To separate the effects of each configuration, rather complex experiments would need to be conducted in order to single out the effects of each property.

EXPERIMENT 2: UNEQUAL INFORMATION SOURCES

In Experiment 2, the component accuracy was varied while the component correlation was held constant for both sources. The goal was to test whether the participants were sensitive to differences in component accuracy between the two sources, and to what extent these differences would affect the participants' weighting strategies, absent any influence from the component correlation. A consequence of this manipulation would be that the information values of the two sources now would be unequal. If participants weighed information from both sources equally, we would be able to conclude that the participants were not sensitive to component (or source) accuracy. If so, the biases observed in Experiment 1 would be attributable to the within-trial consistency differences generated by the manipulation of component correlation. However, if participants assigned larger decision weights to sources with higher component accuracy, we would have to accept the fact that the bias in Experiment 1 was partly due to component accuracy. In that case, we would need to conduct a further experiment to test the effect of component correlation alone.

Method

Participants

The same five participants who took part in Experiment 1 participated in this experiment. They were paid at the same rate as they were in Experiment 1.

Apparatus and stimuli

The same apparatus and stimuli were used in this experiment as in Experiment 1. Displays were generated by the same methods and there was no difference in the appearance of the displays.

Procedure

The experimental conditions are listed in Table 3. There were two major conditions with three subconditions under each. In major Condition I, again labeled the ''small difference'' condition, the d's of Source A's components were 0.96 and were 0.75 for Source B's components. Both sources had the same internal correlation of the components. These correlations were 0, 0.2, and 0.4 for the three subconditions, respectively. Each source's aggregate d' is listed and the ideal ratio of decision weights between the two sources (Source A's/Source B's) is also listed in Table 3. In the three subconditions of Condition I, all sources had the same ideal weight ratio of 1.27. In major Condition II, which we again named the ''large difference'' condition Source A's components had higher detection abilities of 1.11, which caused the ideal weight ratio between Source A and Source B to be 1.48 in each of Condition II's three subconditions. The other manipulations in major Condition II were the same as they were in major Condition I. Each participants finished two blocks of trials with 100 trials in each block in each of the six subconditions. Display position was again counter-balanced as in Experiment 1, and the order of conditions was randomized for each participant. The participants were asked to pay attention to the aforementioned two properties as they were in Experiment 1.

Table 3. Source conditions in Experiment 2: unequal information sources

		Source A			Source B			
Major condition	Subcondition	Component expertise d'_a	Component correlation ρ	Source expertise d'_A	Component expertise d'_b	Component correlation ρ	Source expertise d'_B	Ideal weight ratio d'_A/d'_B
I: small difference	1	0.96	0	1.90	0.75	0	1.50	1.27
	2	0.96	0.2	1.50	0.75	0.2	1.18	1.27
	3	0.96	0.4	1.28	0.75	0.4	1.01	1.27
II: large difference	1	1.11	0	2.23	0.75	0	1.50	1.48
	2	1.11	0.2	1.76	0.75	0.2	1.18	1.48
	3	1.11	0.4	1.50	0.75	0.4	1.01	1.48

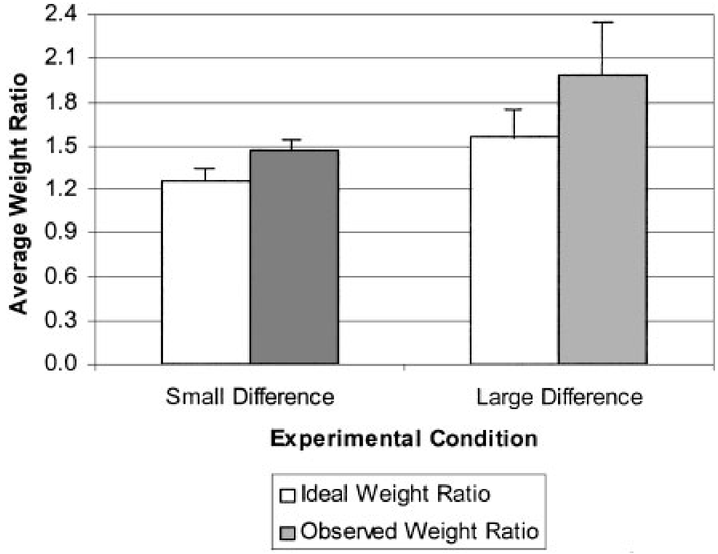

Figure 7. The average obtained and ideal decision weight ratios between Sources A and B in both major conditions of Experiment 2: unequal information sources. The brackets indicate one standard error of the mean.

Results and discussion

We calculated the ideal and observed weight ratios of the two sources, using the same methods as in Experiment 1. Because the ideal weight ratios were the same in the three subconditions of each major condition, we averaged the obtained weight ratios for each participant over the three subconditions within each major condition. From Figure 7, one can see that participants assigned larger decision weights to the source with the higher information value ($t(4) > 5.00$ and $p < 0.01$ when compared to 1 in both major conditions). The mean obtained ratios were significantly larger than the mean ideal ratios ($t(4) = 3.176$, $p < 0.05$, and $t(4) = 2.461$, $p < 0.05$, in the small and large difference conditions, respectively). That is, information from the higher information source had more influence on the participants' final decisions than predicted by the normative model.

Clearly our participants were sensitive to differences in the component accuracy of the two sources. Therefore, the bias observed in Experiment I might be partially attributable to the participants' sensitivity to the accuracy of a source's components. In Experiment 1, the overall information value of the two sources was equal, so the aggregate accuracy of the source having the more accurate components was no better than the aggregate accuracy of the source having components with less expertise. As a result, participants were not able to tell, based on the feedback information, which source's aggregate information was more accurate. This may have forced the participants to find other differences between the two sources to determine an appropriate weighting strategy.

In Experiment 2, trial-by-trial feedback information may have made the difference between sources' accuracy more obvious. Since the source with the higher component expertise was also the source with a higher information value (and was thus more accurate), participants could rely on this source to make their final decisions without any noticeable penalty for doing so. That is, if one source is inferred to have better predictability than the other, the decision maker can ignore opinions from the inferior source. In our experiment, the difference in the achievable percent correct between using a biased weighting strategy (as in Experiment 2) and the ideal weighting strategy would have averaged about 4%. Therefore, a biased strategy could constitute a "fast and frugal" heuristic for participants to achieve reasonably good performance with less effort.

The biases observed in our experiments resemble a phenomenon called ''simultaneous blocking'' found in contingent judgment studies (Chapman & Robbins, 1990; Price & Yates, 1993, 1995). In these studies, participants were required to judge the degree to which the presence/absence of predictor events, such as physical symptoms, was contingent on the presence/absence of an outcome event, such as a disease, after a period of observations. In this situation, participants made predictions about outcomes and received feedback about correctness. It was found that under the same judgment context, the degree of contingency between an outcome event and a moderately contingent predictor event was judged to be stronger when it was accompanied by a weakly contingent predictor event, relative to when the same moderately contingent predictor was accompanied by a strongly contingent predictor event. That is to say, even though the objective contingency of a given predictor event does not change, its subjective contingency is at least partially determined by the other event's objective contingency in the same context. It seems that the relatively stronger predicting cue ''blocks'' the utilization of a relatively weaker cue in these judgment tasks. According to Price and Yates (1993, 1995), this phenomenon can be explained by an associative learning model (Rescorla & Wagner, 1972), which states that the associative strength between the outcome event and the weakly contingent predictor event tends to gain little from trial to trial because the status of the outcome event can be well expected on the basis of the occurrence or nonoccurrence of the strongly contingent predictor event.

The nature of the tasks in those studies is very similar to ours in many aspects. Perhaps the most important similarity is that both tasks can be viewed as a learning process where subjects use multiple cues to make predictions with the aid of outcome feedback. Our result that participants overweighed the source with higher component accuracy as well as higher information value is also explained by the associative learning model: it was hard for the decision makers to recognize the predictability of the source with lower information value when there existed a source with more information value. Although our explanation of the bias has emphasized the economy of different strategies, it should be clear that learning from feedback to distinguish which source's information was a better indicator of the correct decision plays an important role in the explanation of these bias effects.

An additional possibility related to the learning argument is that our participants may have only used the difference between sources' *aggregate* information values to form their weighting strategies, without necessarily noting the specific component differences. This possibility is supported by the way we displayed the two sources' information in the present experiment. The eight-components were separated into two groups with closed rectangles and different colors. According to Gestalt principles of organization for visual objects, it is highly possible that the larger ''groups'' were processed in an earlier stage than the smaller group ''components.'' Actually, the perceptual immediacy of the larger figures over their smaller components is well known and documented in cognitive psychology (Kimchi, 1992). Therefore, information is probably processed at the source level before it is processed at the component level. If a weighting strategy can be determined at the higher level, processing the lower-level information may appear unnecessary.

From the results of Experiment 2 and the above discussion, it is evident that our participants were capable of perceiving and using the accuracy differences between the two sources to form their weighting strategy. Although we speculated that the accuracy differences at the source level were probably the primary cause of the resultant biased weighting, it is hard to exclude the possible effect of accuracy differences at the component level, due to the concurrence of the components' accuracy and the source's accuracy when information correlation is kept constant. However, regardless of which difference in accuracy was primary in strategy formation, both of them originated from the participants' ability to learn from across-trial feedback information. Given that the exactly same kind of feedback was offered in the equal-information-value experiment (Experiment 1), it is possible that the participants could distinguish accuracy differences by using the across-trial feedback. In addition, because the information situation (sources having equal information value) prohibited the participants from discerning differences in source accuracy, we suspect that participants in Experiment 1 might have formed their weighting strategies using component accuracy.

Because of the nature of the confounded relationship between the information value of a source's components, the correlation of the components, and the information value of the source itself, it is impossible to design an independent test of different source information values on participants' weighting strategies. Therefore, in Experiment 3 we chose to test our participants' sensitivity to the component correlation property in a more direct fashion.

EXPERIMENT 3: CONSISTENCY PERCEPTION

Method

Participants

The same five participants who took part in Experiments 1 and 2 participated in this experiment. They were paid the same rate as they were in previous experiments.

Apparatus and stimuli

The same apparatus were used in this experiment as in previous experiments. However, instead of displaying a signal-plus-noise or noise-alone stimulus ahead, the participants only saw the two sources' estimates in each trial. Those estimates were generated by the same methods as in Experiment 1 and had the same appearance as shown in Figure 2.

Procedure

The participants' task in this experiment was to report which source's component estimates were more consistent on each trial. We manipulated the apparent consistency of the four estimates inside each source by manipulating their pair-wise correlations. There were two experimental conditions in this experiment, and all the source and component characteristics were exactly the same as they were in the two conditions of Experiment 1 (see Table 1). Although the components' and sources' expertise levels were irrelevant to the consistency judgment, in order to be consistent with Experiment 1 we kept these manipulations unchanged. Thus, the component correlations were 0.2 and 0 for Sources A and B, respectively, in Condition I; and were 0.4 and 0 in Condition II.

After the display of the two sources' estimates (3000 ms), a pair of "yes" and "no" buttons were displayed. Participants were to click the "yes" button if they thought source A's estimates were more consistent with each other and click the "no" button otherwise. There was no time limit in this procedure. Following their judgment, feedback information was provided to indicate whether the judgment was correct or not. If it was correct, a 2-cent bonus was rewarded; otherwise, participants were penalized 2 cents. Each participant finished four blocks of 100 trials in each experimental condition. As in Experiment 1, the source with the higher component correlation was displayed in the left and marked as "Grp A" in half of the blocks, and in the right and marked as "Grp B" in the other half. The order of conditions was randomized for each participant.

Results and discussion

In order to measure performance in the consistency discrimination tasks, we calculated both the d' and percent correct (PC) of each participant in each experimental condition. When the contrast in correlation was 0.2 vs. 0, the mean d' was 0.29 ($sd = 0.06$), which was significantly greater than zero ($t(4) = 10.544$, $p < 0.001$), and the mean PC was 0.56 ($sd = 0.01$), which was significantly greater than 0.5 ($t(4) = 11.170$, $p < 0.001$). When the contrast was between correlations of 0.4 vs. 0, the mean d' was 0.87 ($sd = 0.11$), also significantly greater than zero ($t(4) = 18.446$, $p < 0.001$), and the mean PC was 0.67

($sd = 0.02$), also significantly greater than 0.5 ($t(4) = 19.699$, $p < 0.001$). These results indicate that our participants were able to discriminate which source's estimates were more consistent in these experimental conditions. Moreover, when the difference in consistency was greater (Condition II vs. I), participants' discrimination improved significantly ($t(4) > 8$, $p < 0.001$ for both d' and PC comparisons).

It is clear from Experiment 3 that participants were sensitive to the within-trial internal consistency differences between the two sources. Since the component correlation of the two sources was the only factor manipulated in Experiment 3, we can conclude that participants would have been able to perceive the correlation differences present in Experiment 1 (because the same correlation levels were tested). Moreover, participants were more sensitive to these differences when the difference in the two sources' internal correlation was increased, congruent with the greater bias found in the large difference condition in Experiment 1. Thus, we can conclude that our participants were able to perceive differences in the internal correlation of the information sources via the consistency judgments.

GENERAL DISCUSSION

This series of experiments demonstrated that when people weigh information from two sources that each consist of multiple components, there is a bias towards the source whose components appear to be more consistent and more accurate, even when such a bias is not normatively appropriate. Experiment 3 demonstrated that a source having a high internal correlation among its components will appear more consistent to the participant, and that people are sensitive to this feature. According to the ideal group model (Sorkin & Dai, 1994), high internal correlation is a property that decreases the information value of a source. If correlation were the basis for a decision bias, the decision maker should prefer sources with low rather than high consistency, because sources with low internal correlation between components are more likely to yield accurate information. This was in fact the result reported by Harkins and Petty (1987). So, why did participants in Experiment 1 of our study favor the high-correlation-high-component-accuracy sources?

First, the experimental situations in the Harkins and Petty (1987) study and the current study were quite different. As pointed out earlier, Harkins and Petty informed their participants about the specific correlation properties of each source before the participants made any response. These explicit statements about the independence of the source's components may have prompted the participants to make correct use of that property. Moreover, participants in the Harkins and Petty study were limited to a single (albeit more complicated) experimental trial, and thus could not form their own trial-based estimates of the statistical properties of the sources. In the present study, participants were told that sources might differ in expertise or correlation, but they were not told which source possessed which property. Furthermore, in this study, the participants were both highly experienced with the experimental task and highly motivated by monetary incentives to maximize their decision performance. Thus, although the results of the Harkins and Petty study disagree with the results of Experiment 1, we would argue that the bias we observed may be a more accurate description of how a human decision maker employs statistical information that s/he has derived from her/his own experience(s) with a source.

The preference for more highly correlated sources is not a unique finding of the present experiment. For example, when studying the relationship between confidence and estimates, Kahneman and Tversky (1973) found that people were more confident about their predictions when those predictions were based on a pair of correlated rather than uncorrelated cues. They concluded that people mistakenly assume that consistency (correlation) implies validity when in lots of cases it does not. In another study, Soll (1999) found that a great portion of his participants valued redundant (correlated) information, even though accuracy was improved more significantly by incorporating nonredundant information. Soll suggested that this might be due to people's different intuitive theories of information. Also, Maines (1996) has argued that when correlation information is offered as a characteristic of information sources, people often either do not know how to make inferences from this information or they will use this information incorrectly. In sum, it seems that human decision makers have trouble estimating a source's actual information value from the correlation property of an information source, and have difficulty using correlated information properly.

An additional factor that may have affected the apparent consistency of the sources is the accuracy of the individual components of the sources. There are two ways in which this statistical property of the source's components could have affected the observer. First, the participants may have observed the greater across-trial accuracy of the individual components of the source that had the higher internal correlation, and then (erroneously) attributed greater overall accuracy to this source, as Experiment 2 suggested. Second, the participant may have been led astray by the partially confounded effects of component correlation and expertise. That is, the estimates from the components of a source may be internally consistent for either or both of two reasons: There may in fact be a high pair-wise correlation among the components of a source (high ρ); or the source components may be independent ($\rho = 0$) but may all possess high expertise (high d_i'). In the latter case, each component would be likely to make a correct decision, and the probability would be high that each component would respond with a similar (correct) estimate on any experimental trial. Without factoring in information about the level of expertise of the components (from across-trial feedback), a decision maker would have difficulty determining which property caused the apparent consistency. In the equal-information-value experiment (Experiment 1), the components of the more highly correlated source also had higher component accuracy, thereby contributing more to this ambiguity.

The above explanations offer some general views about how people's weighing strategies may be affected by their incorrect or incomplete use of sources' statistical properties. Another explanation concerns the participant's use of an averaging process on the source components. In some studies (Clemen & Winkler, 1999; Yaniv, 1997), it has been found that when human decision makers are asked to combine multiple opinions, they tend to average them; in many situations this results in very good performance. As discussed in the unequal-information-value experiment (Experiment 2), given the way our source information was displayed, participants probably processed the information at the source level first before examining the component level. It is very likely that participants employed an averaging process and used the estimated mean ratings of the components inside a source to represent that source's information (just as we did to construct the regression models). As we have noted, a higher pair-wise correlation would result in the component estimates from Source A being more consistent with each other. It is possible that the process of averaging the components of such a source is computationally less demanding than for a source comprised of more diverse estimates. That is, it may have been easier for participants to obtain Source A's aggregate estimate than Source B's, and this computational difference may have facilitated the bias for Source A's information when both sources' aggregate accuracy could not be distinguished. Although this explanation seems to offer a rather specific and alternative view of the bias effect, it is consistent with the idea that the bias occurs as a consequence of the effect of the sources' internal statistical properties on the decision maker's weighing strategy.

Unfortunately, these conjectures leave us unable to separate the precise influence of each property (component accuracy or internal component correlation) on the participants' behavior in Experiment 1. In order to directly measure which property was more crucial in forming the bias, it would have been necessary to titrate both properties so as to find levels of component correlation and component accuracy that resulted in no bias. That is, one could attempt to find a trade-off between the simultaneous effects of the two properties. Having done so, one might then be able to determine which property has the greater influence on the participant's bias.

It is almost common sense that when one is not sure about a decision, one should seek opinions from other people or sources. It is clear from the present study that information from outside sources can be very helpful to a decision maker. With the decision information provided by the two sources in Experiment 1, the improvement in a decision maker's decision performance was impressive, from approximately $d' = 0.9$ to $d' = 2$ overall, producing a 16% increase in accuracy. Although using additional sources information is indeed helpful, integrating information from those sources may be a complicated process.

Refer back to our initial example. After being presented with competing recommendations from the two consulting firms, which one weighs more heavily in your decision? If one firm had a presentation that was excellent and the other firm had an amateurish presentation, then it would be easy to determine which source had a higher information value. You could then weigh the better source more heavily in your decision. However, in this scenario, both firms presented quality proposals, making it difficult to ascertain differences in the aggregate information value of the sources. In such cases where the differences between the groups are not obvious, you may have to judge the quality of the sources based on other information. Experiment 1 demonstrates that as a decision maker, you are sensitive to and may be influenced by both the consistency and expertise of the firm's individual experts. In our scenario, this translates to a bias towards favoring Firm B's advice due to the higher consistency of its experts' opinions (and perhaps their more polished presentation). This over-weighting of opinion from Firm B could be costly, especially if the experts' consistency is due only to prior arrangement and rehearsal rather than the expertise of its judgments. From this example, it should be clear that the process of weighting advice received from different information sources, especially when they disagree, is potentially complex. The statistical properties of sources can have significant effects on how the information they present will be utilized by the decision maker.

ACKNOWLEDGEMENTS

This research was conducted as part of the requirements for the first author's Masters degree. The research was partially supported by a grant from the Air Force Office of Scientific Research to the second author. We are grateful to Frank Yates and three anonymous reviewers for many helpful comments on an earlier version of the manuscript.

REFERENCES

Ashton, A. H. (1986). Combining the judgments of experts: how many and which ones? *Organizational Behavior and Human Decision Process, 38*, 405–414.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37*, 48–74.

Birnbaum, M. H., & Zimmermann, J. M. (1998). Buying and selling prices of investments: configural weight model of interactions predicts violations of joint independence. *Organizational Behavior and Human Decision Processes, 74*, 145–187.

Brehmer, B., & Joyce, C. R. B. (1988). *Human judgment: The SJT view.* New York: Elsevier Science.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460–473.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition, 18*, 537–545.

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis, 19*, 187–203.

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

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grofman, B., Feld, S. L., & Owen, G. (1984). Group size and the performance of a composite group majority: statistical truths and empirical results. *Organizational Behavior and Human Performance, 33*, 350–359.

Harkins, S., & Petty, R. (1981). The multiple source effect in persuasion: the effect of distraction. *Personality and Social Psychology Bulletin, 7*, 627–635.

Harkins, S., & Petty, R. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology, 52*, 260–268.
Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance, 21*, 40–46.
Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.
Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychological Bulletin, 112*, 24–38.
Lee, J. W., & Yates, J. F. (1992). How quantity judgment changes as the number of cues increases: an analytical framework and review. *Psychological Bulletin, 112*, 363–377.
Maines, L. A. (1996). An experimental examination of subjective forecast combination. *International Journal of Forecasting, 12*, 223–233.
Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide.* Cambridge: Cambridge University Press.
Montgomery, D., & Sorkin, R. (1996). Observer sensitivity to element reliability in a multi-element visual display. *Human Factors, 38*, 484–494.
Nitzan, S., & Paroush, J. (1982). Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review, 23*, 289–297.
Price, P. C., & Yates, J. F. (1993). Judgmental overshadowing: further evidence of cue interaction in contingency judgment. *Memory and Cognition, 21*, 561–572.
Price, P. C., & Yates, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 1639–1655.
Richards, V., & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on coefficients. *Journal of the Acoustical Society of America, 95*, 423–434.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64–99). New York: Applenton-Century-Crofts.
Shapley, L., & Grofman, B. (1984). Optimizing group judgmental accuracy. *Public Choice, 43*, 329–343.
Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review, 63*, 129–138.
Slovic, P. (1966). Cue consistency and cue utilization in judgment. *American Journal of Psychology, 79*, 427–434.
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*, 317–349.
Sorkin, R. (1990). Perception of temporal patterns defined by tonal sequences. *Journal of the Acoustical Society of America, 87*, 1695–1701.
Sorkin, R., & Dai, H. (1994). Signal detection analysis of the ideal group. *Organizational Behavior and Human Decision Processes, 60*, 1–13.
Sorkin, R., Hays, C., & West, R. (2001). Signal-detection analysis of group decision making. *Psychological Review, 108*, 183–203.
Tanner, W., & Birdsall, T. (1958). Definition of d' and η as psychological measures. *Journal of the Acoustical Society of America, 30*, 922–928.
Taylor, S. E., & Fiske, S. T. (1978). Salience, attention and attribution: top of the head phenomena. *Advances in Experimental Social Psychology, 11*, 249–288.
Tucker, L. R. (1964). A suggested alternative formulation in the development by Hursch, Hammond and Hursch, and by Hammond, Hursch and Todd. *Psychological Review, 71*, 528–530.
Tversky, A., Slovic, P., & Sattath, S. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*, 371–384.
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.
Yaniv, I. (1997). Weighting and trimming: heuristics for aggregating judgments under uncertainty. *Organizational Behavior and Human Decision Processes, 69*, 237–249.

Authors' biographies:

Shenghua Luan is a graduate student in psychology at the University of Florida. His research interests include applied signal detection theory, information integration in individual and group decision making, and information consulting and purchasing.

Robert D. Sorkin is a Professor in the Departments of Psychology and Industrial & Systems Engineering at the University of Florida. His recent research has concerned the development of signal detection models of group behavior.

Jesse Itzkowitz is a graduate student in the cognitive psychology program at the University of Florida, where he also received his undergraduate degree. His research interests include group decision making, signal detection theory, and computational modeling. His current project involves the examining and modeling the effects of group composition and organizational structure on performance, bias, and decision time.

Authors' address:
Shenghua Luan, Robert D. Sorkin, and **Jesse Itzkowitz**, University of Florida, Gainesville, FL 32611-2250, USA.