**Title and Subtitle:** Assessing and Improving Team Decision Making

**Author(s):** Robert D. Sorkin

**Performing Organization Name(s) and Address(es):**
University of Florida
P.O. Box 112250
Gainesville, FL 32611-2250

**Performing Organization Report Number:** N/A

**Sponsoring/Monitoring Agency Name(s) and Address(es):**
Air Force Office of Scientific Research (AFOSR)
801 North Randolph Street Room 732
Arlington, VA 22203-1977

**Supplementary Notes:**

**Distribution Availability Statement:**

**Abstract (Maximum 200 words):**

This project employed analytical and experimental techniques from signal detection theory to (a) assess the accuracy of team performance, (b) identify sources of inefficiency in team decision making, (c) specify how team members utilize information received from sources having different statistical properties, and (d) model the team deliberation process. The team's task was to decide whether signal-plus-noise or noise-alone had occurred, based on individual graphical displays presented to the team members. Team performance was shown to depend on (1) the signal-to-noise level of members' displays and the efficiency of individual member detection (compared to the statistical optimal), (2) the correlation (common noise) among members' displays, (3) constraints on member communication and interaction, and (4) how efficiently the team combined member judgments to form the team's decision (including mandatory voting rules). The internal correlation, expertise, and bias of different information sources influenced the decision weights that team members gave to these sources, sometimes in a non-optimal fashion. Using a Bayesian network model of team deliberation, the project began to quantify important aspects of the deliberation process such as the interaction protocols used by members of a networked team.
1. Cover Page/Abstract (use Form OMB 0704-0188)

ABSTRACT

This project employed analytical and experimental techniques from signal detection theory to (a) assess the accuracy of team performance, (b) identify sources of inefficiency in team decision making, (c) specify how team members utilize information received from sources having different statistical properties, and (d) model the team deliberation process. The team's task was to decide whether signal-plus-noise or noise-alone had occurred, based on individual graphical displays presented to the team members. Team performance was shown to depend on (1) the signal-to-noise level of members' displays and the efficiency of individual member detection (compared to the statistical optimal), (2) the correlation (common noise) among members' displays, (3) the constraints on member communication and interaction, and (4) how efficiently the team combined member judgments to form the team's decision (including mandatory voting rules). The internal correlation, expertise, and bias of different information sources influenced the decision weights that team members gave to these sources, sometimes in a non-optimal fashion. Using a Bayesian network model of team deliberation, the project has begun to quantify important aspects of the deliberation process such as the interaction protocols used by members of a networked team.
FINAL PERFORMANCE REPORT

Assessing and Improving Team Decision Making

AFOSR/NL
Grant F49620-00-1-0041

February 1, 2002

Robert D. Sorkin
University of Florida
Gainesville FL 32611-2250
1. EXECUTIVE SUMMARY

This project employed signal-detection theory to study the performance of human decision making teams. The project's goals were to (1) quantify the decision making performance of teams, (2) identify sources of inefficiency in team decision making, (3) specify how information received from sources having different statistical properties is weighed and combined, and (4) predict the duration and other important characteristics of the deliberation process. The team task was to decide on the presence or absence of signals presented in noise. These signals were presented to team members on individual graphical displays and the team arrived at a decision based on partial or full communication among team members.

An important goal was to define and control the key factors that affect team performance in the laboratory, including: member expertise (which was manipulated by controlling the signal-to-noise ratio of the members' displays); member response bias (controlled via the prior odds of signal and the pay-offs for correct and incorrect decisions); correlation among member judgments (controlled via the level of common display noise); and the imposed decision rule, team structure, and constraints on member intercommunication. Member expertise was defined as each member's ability, $d'_i$, to accurately discriminate among the decision alternatives, and member bias was defined as a member's tendency, $c_i$, to favor one decision alternative over another based on other than the input evidence. The project studied several types of member interaction and decision rules, including: (a) no information received from other members; (b) binary (yes/no), continuous (graded), or verbal responses from members; (c) information or no-information provided about the expertise and/or bias of other members; (d) one-time vs. iterated communication of member responses; (e) fixed majority rule (1/2, 2/3, 3/4, or unanimous); and (f) consensus or hierarchical decision structure. In addition, the project defined three quantifiable aspects of team member diversity important to predicting team performance: (1) the variance in team member expertise $\sigma^2_{d'_r}$, (2) the variance in member bias $\sigma^2_c$, and (3) the correlation $\rho_{i,j}$ between the judgments of any pair of team members.

The performance of a statistically optimal team is given by the equation:

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

where $d'_{\text{ideal-group}}$ is the detection index that specifies the accuracy of the team's performance, $m$ is the number of members, $\rho$ is the correlation between a pair of members, and $\mu_{d'_r}$ and $\sigma^2_{d'_r}$ are the mean and variance, respectively, of the set of detection indices $\{d'_i\}$ that characterize the members of the team (Sorkin and Dai, 1994). Figure 1 shows a diagram of such an ideal distributed detection system. Values of $d'$ equal to 1 and 2 correspond, respectively, to percent correct performance of approximately 69% and 84%.) The equation shows that optimal detection accuracy will increase with $\sqrt{m}$ and will
decrease as ρ differs from zero. The individual member outputs should be weighted by the set of weights \( \{d'_i\} \), where \( d'_i = [1 + \rho(m - 1)d'_i]^{-1} - m\rho \mu_d' \).

![Diagram of an ideal group signal-detection system composed of \( m \) members. Each member is subjected to two sources of Gaussian noise: one unique (\( \sigma_i^2 \)) and one common (\( \sigma_{\text{common}}^2 \)) to the other members. The decision variable, \( Z \), is formed from the weighted sum of the continuous member estimates.

The project assessed the performance of teams & members (with teams of from 5 to 15 persons) over thousands of detection trials in a graphical signal detection task. The potential source of inefficiencies at each stage of decision making was evaluated by comparing team and member performance to the optimal statistical prediction. In general, the effect of the experimental variables on performance was consistent with the ideal detection analyses, and in some conditions the performance of the human teams approached the statistical optimal. That is, the squared ratio of the obtained detection index to the ideal detection index was equal to or greater than 0.8. In particular, (1) teams were highly efficient at aggregating the information received from the team members and (2) member responses were appropriately weighted by knowledge of the member’s expertise and bias. The latter effect was specifically tested in a condition (conformity situation) in which the binary judgments of some members were biased by experimentally manipulating their payoffs. In this condition, members appropriately de-rated the biased inputs and employed a rational aggregation strategy.

In some conditions performance deviated from the optimal prediction. For example, in one set of conditions individual detection effort decreased as a function of group size. This strategy may be interpreted as an efficient way to maintain attention to
secondary aspects of the main team task. In a second set of conditions, members were asked to base their decisions on combinations of information sources. The statistical correlations of these sources were manipulated so as to vary the internal consistency of the sources' judgments. Members exhibited an inappropriate preference for (i.e., gave higher decision weights to) information sources that had higher component detection indices and higher internal consistency, which resulted in a decrease in detection efficiency and performance. This decision bias has not been previously reported in the literature.

The project also began to study the behavior of teams engaged in a deliberation process in which the judgmental estimates from each member are communicated in a sequential rather than parallel manner. During this process, members use a Bayesian rule to update their current signal estimate based on the responses of other members and on knowledge of the other members' expertise and bias. One goal is to predict how aspects of the deliberation process, such as duration and accuracy, depend on the specified member characteristics and on the protocol that determines the order of member responding. A second goal is to discover the response protocol(s) actually used by team members in unsupervised distributed detection networks and to develop optimal protocols for networked teams. This component of the project is in its early stages.

2. ACCOMPLISHMENTS/NEW FINDINGS

Efficiency of Team Decision Making

A major accomplishment was a research study that successfully addressed the following questions: How effectively can teams of people make yes-or-no decisions and how does their performance depend on the abilities of the individual members and the way they interact? Signal detection theory was used to model the behavior of teams of human participants in a visual detection task. The model specifies quantitatively how performance depends on the team's size, the competence of the members, the nature (binary or graded) of and the correlation among members' judgments, the constraints on member interaction, and the team decision rule. The model quantifies the effect on performance of using non-optimal rules for team decision, such as the arbitrary weighting of member judgments, the uniform weighting of judgments ("Delphi" groups), and the use of binary (yes-no) or continuous (graded) member voting. The model enables the specification of team performance efficiency, which is a measure of how much a team's performance differs from the statistically optimal group. The article showed how this efficiency measure can be factored into separate components that describe (a) how well the team members performed their individual detection tasks and (b) how effectively the team combined the information received from the members into a group decision. The performance of the studied teams was consistent with the theoretical predictions, but efficiency decreased as team size increased. This result was attributable to a small decrease in the effort that members allocated to their individual tasks rather than to inefficiencies in combining the information in the members' judgments. This is perhaps the first definitive study showing that (a) it is possible to make quantitative predictions
about the level of team performance based on the known levels (and correlation) of member expertise, and (b) that teams are capable of aggregating judgments received from the team's members in a near optimal fashion. (Sorkin, R.D., Hays, C.J., and R. West. (2001). Signal detection analysis of group decision making. \textit{Psychological Review}, 108, 183-203.)

\textbf{Detection in Simulated Teams}

The group decision efficiency experiments (described above) were extended to include teams with simulated members. Each observer was placed in an individual decision task in which from 3 to 17 simulated team members provided uncorrelated estimates of signal occurrence. In contrast with the results of the experiments with human teams, there was no evidence of a decrease in either the observer's individual detection effort or in the efficiency of information aggregation of the (simulated) member estimates as a function of team size. The decision weights that the observers assigned to their own estimates and the estimates of the simulated members closely approximated the optimal ideal weights for these situations. The disagreement between this result and the earlier study may be due to unspecified social factors in the first study that involved the physical presence of team members, such as the perception of a shared responsibility for team performance.

An explanation of the difference between the experiments is suggested by an analysis of team performance and member detection efficiency. This analysis indicated that a small decrease in the detection expertise (d') of all the team members (increasing with team size) results in small drop in team percent correct; this drop increases slowly as team size is increased. For a 100-trial block and a 15-member team, the drop in percent correct would be very difficult to observe (approximately the size of the standard error.) However, the corresponding drop in an individual member's (d')² increases dramatically with increases in team size. Thus it is possible for each team member to achieve a finite decrease in individual effort with only a small decrease in the team's performance. The (d')² is appropriate as a measure of detection effort because it is directly related to the physical energy of the signal and thus the observer's detection effort, and because dual-task time-sharing experiments usually show a curvilinear trade between the d' score on each task. That is, the performance of a subject is summarized by the sum of the (d')² on the two simultaneous tasks being equal to a constant. Thus a small decrease in d' on one task can yield a small but practical level of performance on a second, simultaneous task. The small decrease in effort on the primary task may be useful to a team because it can allow attention to other events occurring in the team's environment. The simulated team experiment allowed far fewer distractions to draw attention away from the primary detection task than did the human team experiment. This is consistent with the lack of a decrease in individual member detection performance in that task. This hypothesis could be examined in future studies by requiring team members to simultaneously perform a secondary task along with the team task, and measuring performance on both tasks. (Sorkin, R.D., Luan, S. and Itzkowitz, J. Rational Models of Social Conformity and Social Loafing. European Association for Decision Making (Subjective Probability, Utility, & Decision Making-18) Meeting, Amsterdam, August 2001.)
Knowledge of Member Bias and Expertise

People usually make decisions that conform to the known opinions of the people around them. A traditional argument says that this bias toward the majority view is caused by social pressure in the decision maker’s environment. An alternative argument is that conforming behavior exists because the conforming choice is almost always the rational choice. That is, agreeing with the opinions of other team members is likely to result in decisions that are accurate and therefore will have positive outcomes for the decision maker and the team. This experiment tested conditions in which the rational and conforming strategies prescribed different behavior. Small teams of observers performed a graphical decision task. Each team member observed a signal-plus-noise or noise-alone input and then made an individual yes/no vote. After these votes were displayed to everyone, one member of the team was randomly selected to make the team’s final decision. Monetary payoffs were based both on the accuracy of the first votes and the accuracy of the team decision. Two quantitative models of the team decision were evaluated: (a) a rational strategy that incorporated the first votes into an ideal detection algorithm and (b) a conforming strategy that simply complied with the majority outcome of the first vote. Both strategies predicted similar performance when the first vote was taken under equal penalties for members’ misses and false alarms. However, these strategies predicted different behavior when the first vote was biased toward a yes or a no response (i.e., a rational decision maker should discount information known to be biased). The results indicated that team members used the rational rather than the conforming strategy. This indicates that the rational motive is dominant when actual payoffs are involved. The result shows that decision maker knowledge of member bias can be used to appropriately discount such inputs in decisions. (Sorkin, R. D., West, R., & Luan, S. Social influence in group signal detection. 41st Annual Meeting of the Psychonomic Society, New Orleans, LA, November 2000; an article describing these results is in preparation for the Journal of Behavioral Decision Making.)

These experiments were extended to cases where both the bias and the expertise of the team members were varied during the experiment and when the decision maker was or was not given this information. The goal was to further quantify the conditions under which human teams can optimally aggregate information from multiple information sources. This is a situation that is common to many military and civilian decision environments; the sources may be other people or may be machine systems or sensors. The results indicated that rational decision strategy was the best predictor of team member behavior. (Sorkin, R.D., Luan, S. and Itzkowitz, J., Rational Models of Social Conformity and Social loafing, European Association for Decision Making; Subjective Probability, Utility, & Decision Making-18 Meeting, Amsterdam, August 2001; and Sorkin, R. D., Luan, S., Itzkowitz, J., and West, R., The Informational Value of Group Interaction, 42nd Annual Meeting of the Psychonomic Society, Orlando, Florida, November 2001.)
Effect of Consistency and Correlation of Information Sources

People often make decisions using summary advice from other sources. For example, a person faced with a medical or a financial decision will often obtain opinions from other persons around them. Likewise, a decision-making member of a military team may obtain additional estimates of the situation from other team members, automated sensors, or other teams. Some sources of information may be more reliable than others, composed of more sub-elements, and these sub-elements may be more internally consistent or partially correlated with the judgment of the person desiring the information. This study asked how decision makers weight the estimates received from different sources when those sources vary in their (a) number, (b) reliability or expertise, (c) apparent consistency (internal correlation), and (d) correlation with the decision maker's initial estimate. These experiments involved human decision makers and simulated team members in a graphical decision task.

As in the project's other experiments, the decision maker in these tasks observed a display of a signal-plus-noise or noise-alone event and then made an estimate of the signal's likelihood. The decision maker was then supplied with a graphical display of several estimates made by one or two sets of simulated team members. The decision maker was then required to make a final yes-no decision about the occurrence of signal on that trial. The payoff to the decision maker was based on the accuracy of the final yes-no decision. Performance was assessed when the estimates of the virtual team members were pair-wise uncorrelated or correlated (among themselves) at some level, or had differing overall indices of detection. By computing the point-by-serial correlation over trials between the decision and the source's magnitude, the experiment assessed how much decision weight the decision maker gave to each information source.

One experimental condition tested which of two equal-information sources (sources with equal aggregate-d') would be given the higher weight: the one with the internal pair-wise correlation and higher individual d's, or the one with the zero pair-wise correlation and lower component d's. Thus, this condition tested for the presence of a bias toward sub-source consistency or sub-source expertise. Further conditions were run comparing sources with differing overall informational value. The results of these experiments indicated that there is a small but highly significant bias toward information sources that have higher consistency and higher component expertise, even though the information available from such sources is identical to or less than that received from lower consistency sources. This is an important result because it shows the existence of a here-to-fore unreported bias toward information sources that have higher component expertise or internal consistency. Such a bias will result in inefficiency in decision performance; performance will decrease as a function of the discrepancy between actual and optimal weighting of input sources. It may be possible to provide a decision maker with the estimates received from different sources so that the overall information display adjusts for such pre-existing weighting biases. (Luan, S. and Sorkin, R. D., Weighing information from different sources: Are people rational? Annual Meeting of the Society for Judgment and Decision Making, Orlando, Florida, November 2001; Luan, S. Utilization of information from outside sources in decision making., Masters Thesis,
University of Florida, 2002. An article summarizing these experiments is being prepared for the *Journal of Behavioral Decision Making*.

**Deliberation Networks**

The project began to model the process by which the members of a team share their individual estimates and deliberate until a decision is reached. For example, there may be an imposed criterion for member consensus such as a three-quarter or unanimous majority and deliberation would continue until that criterion of agreement (or some time limit) is reached. The manner of member interaction may be face-to-face or constrained via a computer or communications network. One project goal is to predict how the duration and dynamics of the deliberation process depend on the members' competencies, between-member correlations, interaction constraints, and decision rules. Other important goals are to determine how performance is affected by the protocol that underlies the members' response sequence, which common response protocols are used by members, and whether it is possible to implement an optimum protocol.

A basic assumption is that team members employ a rational strategy for updating their response criteria using information from the other members' responses. This rule has been applied by authors in the context of human signal detection (cf. Murrell, 1977; Robinson and Sorkin, 1985), the design of distributed detection systems (e.g., Papastavrou, 1992), and so-called "information cascades" in economic behavior (Anderson and Holt, 1997; Huck and Oechssler, 2000). In our simplification of the team decision situation, members update their decision criterion by revising their prior odds ratio estimate using a Bayesian rule to integrate the other members' expressed judgments and information about member d's, and criteria. That is, each time any member "speaks," every other member revises his/her estimate of the posterior odds by computing a new product of likelihood ratio and prior odds ratio. After each update, members reform their positions and a test is taken for consensus. This process is schematized in figure 2.

It can be shown that the accuracy of the group's final decision is dependent on the sequence of responses made during the deliberation process. The project has begun to evaluate the effect (and existence) of different rules controlling the likelihood that a member will speak at any point in the member response sequence. For example, one rule is that the probability that a member speaks is a function of the absolute difference between that member's likelihood estimate and current response criterion. Other rules include consider that the member with (a) the highest magnitude observation, (b) highest expertise, and (c) the most extreme criterion, will be most likely to speak. Whatever the rule, the influence of any new response (or discussion) will be small after several iterations of this process. The deliberation process may be terminated by a decision rule that requires a specific degree of consensus, a sufficiently small change in the expressed opinion, or the attainment of a fixed time limit or number of cycles. The project is investigating the effects of different response protocols and different stopping rules on simulated and human teams.
Observe input and calculate likelihood ratio:

\[ \beta = \frac{(V_{\text{correct-no}} + V_{\text{false-alarm}})}{(V_{\text{hit}} + V_{\text{miss}})} \cdot \frac{p(n)}{p(sn)} \]

Set initial criterion

\[ I(x) \]

\[ I(x) > \beta ? \]

Make initial vote (yes, no)

Obtain initial votes from other members \( \{y_1, n_2, n_4\} \) and recalculate response criterion:

\[ \beta_F = \frac{(V_{\text{correct-no}} + V_{\text{false-alarm}})}{(V_{\text{hit}} + V_{\text{miss}})} \cdot \frac{p(n \mid y_1, n_2, n_4)}{p(sn \mid y_1, n_2, n_4)} \]

\[ I(x) > \beta_F ? \]

Make final vote (yes, no)

Aggregate members' final votes for decision.

Information about members' expertise and criteria

\( d^{-1}, d^{-2}, d^{-4}, \beta_1, \beta_2, \beta_4 \)

Figure 2. Schematic of distributed detection system for binary-responding members.

Although it is early in the project's progress on this problem, the simulations indicate that the group decision usually converges to the initial majority opinion and that performance is dependent on the sequence-determining response protocol. The latter is a potentially important result because it suggests that one may be able to devise an optimum protocol for decision networks that must make binary decisions based on the individual judgments of the network members.
References Cited


3. PERSONNEL

Robert D. Sorkin
Professor, Department of Psychology
University of Florida

Christopher J. Hays
Graduate Student, Department of Psychology
University of Florida (now serving in the US Air Force)

Ryan West
Graduate Student (ABD), Department of Psychology
University of Florida (Now at Microsoft Corporation, Seattle, WA.)

Shenghua Luan
Graduate Student, Department of Psychology
University of Florida

Jesse Itzkowitz
Graduate Student, Department of Psychology
University of Florida

4. PUBLICATIONS & THESES


Pending submission:


5. PAPERS AT PROFESSIONAL MEETINGS


6. INTERACTIONS & CONSULTATIONS


Member, Editorial Board of Human Factors

7. DISCOVERIES/INVENTIONS None

8. HONORS/AWARDS None