Journal of Personality and Social Psychology
1993, Vol. 65, No. 5, 959-974

0022-3514/93/$3.00

The Common Knowledge Effect: Information Sharing and Group Judgment

Daniel Gigone and Reid Hastie

The hypothesis that the influence of an item of information on a group judgment is directly related to the number of group members who hold that information before group discussion was tested. Three-member groups read short descriptions of students and were asked to make individual and then group consensus judgments about those students' grades in the course. Information held by all members before group discussion had more influence on the group judgments than information held by only 1 member. However, no effect of information distribution was found when controlling for member judgments, suggesting that the impact of the information, and hence the effects of distribution across members, was mediated by its impact on individual-member prediscussion judgments. The group judgments were no more accurate than the average of the member judgments. Group members were not aware of the common knowledge effect's influence on their use of information.

The exchange and sharing of information is an important process in decision-making groups. The ability of groups to consider more information and to consider information from diverse sources is the primary reason that groups are expected to make better decisions than individuals acting alone. Members bring different informational resources to the group, which can be pooled to produce a group decision that is of higher quality than any of the group members could have produced with less complete information. Decision-making committees are often brought together for the express purpose of making a more informed decision through the exchange of initially unshared information. Committee members are often chosen with different backgrounds so that the committee will be able to make decisions based on a wider range of information. However, if groups do not pool information effectively, they will not make better decisions than individual decision makers.

The exchange of information in a group judgment task is essential when the various items of information are not initially available to all group members. In many group judgment situations, as when committee members are chosen from different backgrounds, individual group members are likely to bring different items of information to the group. In such a case, each group member can only base his or her opinion on the information that he or she brings to the task and any information that is exchanged with the group by other group members. If items of information are not exchanged, the group judgment will be based on incomplete information.

Daniel Gigone and Reid Hastie, Department of Psychology, University of Colorado.

We thank Charles Judd and David Kenny for their advice concerning data analysis and Garold Stasser, John Rohrbaugh, and Tom Stewart for their suggestions on the research plan. Support for this research was provided by National Science Foundation Grant BNS-8996145.

Correspondence concerning this article should be addressed to Daniel Gigone, Department of Psychology, University of Colorado, Boulder, Colorado 80309.

Stasser and his colleagues (Stasser, Taylor, & Hanna, 1989; Stasser & Titus, 1985, 1987) have invented a useful method to study the effects of information distributions across group members under controlled laboratory conditions. They gave subjects information about hypothetical candidates for the position of student body president. Some items of information were shared by all of the group members and some were unshared, held by only one of the group members. In some experimental conditions, 66% of the information about each candidate was fully shared, whereas in other conditions, only 33% of the information about each candidate was shared. Groups were then convened and asked to make a consensus decision on which of the candidates was most qualified for the position. The basic paradigm has been repeated with different-sized groups, with different total amounts of information about each of the hypothetical candidates, with different proportions of the information shared and unshared, with structured and unstructured group discussions, and with the overall available information biased toward one of the three candidates or the available information distributed to make the candidates equally attractive. Dependent measures have included the candidate selected, pre- and postdiscussion recall of informational items, pre- and postdiscussion individual judgments, and observational measures of the items of information that were exchanged during group discussion.

Stasser's studies showed that subjects generally do not effectively pool unshared information. Groups are more likely to discuss shared information than unshared information. This tendency increased with larger groups, higher overall information load, and a higher percentage of unshared information. Moreover, the tendency to discuss shared information biased the final group judgment in the direction of the initial information distribution. This bias occurred when the available information, if fully exchanged, should have led to a judgment in favor of a different candidate than the one who was initially favored. Only when there was a low overall information load and a high percentage of unshared information was there substantial discussion of the unshared information, and even here there was still a substantial bias in group decisions caused by the initial information distribution pattern.

Stasser explained his results in terms of an information sampling model (Stasser & Titus, 1985). He claimed that the discussion bias toward shared information can be explained by looking at the probability that any one item of information will be recalled by any one of the group members. The task is disjunctive, because only one group member needs to remember an item in order for it to be discussed by the group. Thus, there is a higher overall probability that a shared item will be recalled by one of the group members than that an unshared item will be recalled by the only group member who possesses it. Stasser was able to use his information sampling model to explain his results concerning group size, information load, and level of information sharing.

However, Stasser's information sampling model fails to tell the whole story about information sharing and group judgment. It assumes that group judgments are the result of a quantitative summation of evidence for and against each candidate, that the group decision process is dominated by "informational influence" or "persuasive arguments" (cf. Burnstein, 1982; Deutsch & Gerard, 1955). His model, because it explains biased judgments in terms of the simple probability of discussion of an item of information, weights all items of information equally in the determination of the group judgment. Moreover, because the model focuses exclusively on the discussion of information, it does not necessarily account for all of the mechanisms that could lead to differences in influence between shared and unshared information. For example, unshared information that contradicts the prevailing opinions might be seen as more surprising and informative than unshared information that is in accordance with the prevailing opinions, and thus it may have a greater impact on the group discussion and its outcome (cf. Hastie, 1984). Also, information that is used to form an individual judgment might be weighted differently than information that is learned later through group discussion.

The present study was therefore designed to look more closely at the relationship between information sharing and the influence of information on group judgments. Specifically, it was designed to test the hypothesis that an item of information will have more influence on the judgment of a group when it is shared than when it is unshared. What we will call the "common knowledge effect" hypothesis can be stated more precisely: "The influence of a particular item of information is directly and positively related to the number of group members who have knowledge of that item before the group discussion and judgment."

Intuition suggests two ways in which shared information could have a relatively large influence on group judgments. First, information held by group members will likely influence their individual judgments. Those judgments, in turn, will influence the group judgment. A piece of information that is held by several group members will affect all of those members' prediscussion judgments and will therefore have a larger influence on the group judgment simply because of its impact on the members' prediscussion judgments. For example, information from a job candidate's job talk will influence the judgments that the audience members form about that candidate. When the hiring committee meets to make a group judgment about that candidate, the committee members probably will express their own judgments, which were influenced by the job talk. Even if the job talk itself is not mentioned, its impact will be conveyed through the members' opinions.

Second, as Stasser has emphasized, shared information can have an undue influence on the group judgment because it is a common reference point for the group members. Shared information might then be more likely to be discussed and evaluated during group discussion. The hiring committee might spend more time discussing members' impressions of the job talk, which all group members have attended, than the candidate's publications, which only a few members have read. If the group discussion focuses on shared information, that information could have inordinate influence on the group judgment. One goal of the present study is to distinguish between these two possible mechanisms that could mediate the hypothesized common knowledge effect.

In the present study, groups made judgments about the grades received by other undergraduate students in an introductory psychology class. These judgments were based on information about each to-be-judged student's academic ability (e.g., high school grade point average and aptitude test percentile score) and motivation (e.g., enjoyment of the course). Obviously, the given information was relevant to the judgments that were to be made. However, there was likely to be disagreement among group members concerning the importance of the various items of information and the relationship of each item to the course grade of a typical psychology student. The grade judgment task was chosen for two reasons. First, undergraduate subjects should be familiar with making this type of a judgment. They undoubtedly have the opportunity to infer from uncertain information what grade they themselves or other students will receive in a course. Second, the grade judgments can be tested for accuracy. Judgments about a stimulus case can be compared with a criterion, the actual grade that the stimulus student received in the course.

The present study used a task that was designed to maximize the probability that unshared items of information would be discussed by the group. Subjects were assigned to three-member groups. Moreover, the total available information about each to-be-judged student consisted of only six items. In the present study, two of the information items were fully shared, two items were unshared, and two items were known by two of the three group members before any discussion. In addition, subjects made a series of 32 judgments, receiving as a group the same six items of information about each to-be-judged student. Thus, each group member should have been aware of the types of information that were available to the group and which of those he or she did not receive. According to Stasser, all of these factors should have led to a high level of discussion of unshared information. The groups in the present study were the same size as Stasser's smallest groups. Even in the lowest load condition, each of Stasser's subjects had to remember 12 items of information about each of three hypothetical candidates, for a total of 36 items. The present study does not correspond exactly to any of Stasser's levels of sharing, although it could be seen as falling somewhere between the 33% and 66% conditions. However, the common knowledge effect hypothesis predicts that shared information should have more influence on the group judgment, even if the unshared information is discussed fully.

Although the present judgment paradigm has some valuable qualities, questions can be raised concerning its external validity. Artificial groups of subjects made a large number of judgments in a relatively short period of time and at low cost to themselves or others. These conditions might have led subjects to use simple heuristics in making their member and group judgments. Moreover, each group member made a written judgment before the group discussion of each case. Thus, a member might be likely to be committed to his or her initial judgment. Such a commitment effect might discourage members from discussing or considering additional information while they are attempting to reach a group consensus. In sum, the results of the present study would not be of particular interest if they were primarily due to the structure of the research paradigm and if that paradigm did not correspond to common real-world group judgment settings.

However, the present paradigm was analogous to a number of authentic group judgment settings. Many committees meet to make several similar judgments or decisions from a relatively small amount of information. For example, a university petition review committee may make a series of quick judgments concerning the merit of various petitions to waive requirements, extend deadlines, and so on; a college admissions committee will consider the strengths and weaknesses of hundreds of applicants; a salary review committee will meet periodically to assess the amount of salary increase deserved by various employees; and a research grant review panel will meet to review 50 proposals. In each of these cases, the group has limited time in which to make a large number of decisions. In addition, members of such groups might implicitly, or even explicitly, commit themselves to individual judgments before group discussion of each case. For example, each member of the admissions committee might make a recommendation based on his or her individual review of an applicant's file. Only if there is disagreement about a particular applicant might group discussion of that case occur. Moreover, in any of the above examples, judgment-relevant information could be differentially distributed among group members. We contend, therefore, that the present research paradigm is relevant and general.

A recent study by Stasser and Stewart (1992) provides some insight into what might be expected from the present judgment task. They found that the bias resulting from unshared information was greater for judgmental tasks, for which groups make judgments based on an assessment of evidence, than for intellective tasks, for which groups believe that there is a demonstrably correct solution to a problem. Groups that do not believe there is enough information to reach a demonstrably correct answer will be motivated in their information search to reach a consensus. The present judgment task clearly falls on the judgmental end of the task continuum. Although our subjects were informed that there was a "correct" judgment for each case—the actual grade received by that student—they had no way of demonstrating the correctness of their judgments. Thus, according to Stasser and Stewart, the unequal distribution of information would be likely to bias group judgments in our research task. Likewise, differential distribution of information among group members might bias the judgments of real-world groups such as the ones described above. However, Stasser and Stewart again explained the differences between types of tasks in terms of the tendency of group members to be biased in their sampling of unshared information from memory. We argue that judgment bias in the present task cannot be explained by differential pooling rates for shared and unshared cues.

In his review of group accuracy research, Hastie (1986) made a similar distinction between types of group judgment tasks. He argued that groups tended to be more accurate than their average members only when the task involved a demonstrably correct solution. In the present task, then, a group's judgments would not be expected to be more accurate, on average, than the mean of its members' judgments, at least in the fully shared information control condition. However, with differential distribution of information among group members, a group could have the potential to judge more accurately than the average of its members. Valuable information that was withheld from some group members before their member judgments could be pooled during the group discussion. Thus, whereas the members would typically make many of their judgments without the aid of some judgment-relevant information, the group would be potentially fully informed before every judgment. Such groups might be expected to make more accurate judgments than the average individual group member.

Social judgment theory (Brehmer & Joyce, 1988; Hammond, Stewart, Brehmer, & Steinmann, 1986) is a framework for the analysis of the use of information in judgment. An item of information, which is called a *cue*, has some true relationship with the *criterion*, the true state of the judgment target. There is, for example, some true relationship between the high school grade point average of a sample of students and the grades that they tend to receive in introductory psychology courses. The relationship between cues and the criterion can vary in both strength ("ecological validity") and form. For example, a cue could be more or less strongly related to the criterion and that relationship could be either linear or quadratic. Symmetrically, there is some relationship between each cue and the judgment that is made, and these relationships can vary in strength ("cue utilization") and form.

Because of the symmetric relationships between cues and criterion and between cues and judgment, this formulation is known as the "lens model" of judgment (Brunswik, 1956; Hammond, 1966). The strength and form of the relationships between cues and judgments across multiple judgments is known as the judgment policy of the judge. To the extent that this judgment policy matches the true relationships between criterion and cues, the judgments will be accurate, differing from the criterion only because of random error. In making a judgment, then, the individual or group can be described as using a judgment policy to infer the criterion from the available set of cues.

Across a series of judgments, the judgment policy of an individual or group can be determined by running a regression analysis, with the set of cues used in the task as the independent variables and the judgment that is made as the dependent variable (Hammond et al., 1986; Harmon & Rohrbaugh, 1990; Slovic & Lichtenstein, 1971; Stewart, 1988). The presence of reliable interaction or higher-order coefficients would indicate that there were dependencies between cues or that the form of the cue utilization was nonlinear. Within a judgment task, the relative importance or influence of the cues can be compared by standardizing the regression coefficients to account for the scaling and variation of the cue values. Thus, social judgment theory provides a method for comparing the influence of a cue on judgments under different experimental conditions. Although this statistical policy-capture approach has been used extensively to study the influence of information on individual judgments, it has not, to our knowledge, been used before to study group judgment policies. Other researchers have measured changes in the judgment policies of individuals resulting from the resolution of group judgment conflict (cf. Brehmer & Hammond, 1973).

Thus, the group judgment policy provides one useful summary of the group process. Variations in the group's information usage in making a judgment are reflected in the policy weights that correspond to the different cues. A cue that has a strong influence on the judgments made by the group will be heavily weighted in the group judgment policy. The issue of what affects the utilization of information in group judgment can therefore be restated in terms of the group judgment policy. Those factors that affect the cue utilization weights in the group judgment policy do so through their influence on the group's usage of information in making judgments.

Several factors will affect individual judgment policy weights in a multiple cue judgment task. First, the judgment policy will be influenced by the judge's "theories" about the usefulness of each cue. If a person believes that aptitude test scores are the best predictors of college academic performance, that cue will be weighted heavily in the person's judgments about college course grades. Second, judgments will be influenced by cue–response scale commensurability (Slovic, Griffin, & Tversky, 1990; Slovic & MacPhillamy, 1974). A cue that is expressed on the same scale as the required judgment response is likely to receive more weight than a cue that is more difficult to map onto the response scale. Third, cue order, either in presentation or in discussion, probably affects judgment policy weight. A cue that is always presented first or discussed first might anchor a person's judgments, with later cues leading to adjustments in the judgments. Fourth, the salience or dramatic quality of a cue might affect its weight in judgment. A cue that is presented in a novel typeface or that is likely to evoke a vivid image or an emotional reaction, for example, might have more weight than if it were presented in a less distinctive manner. Finally, we expect that the weight of a cue in the group judgment policy will be influenced by its distribution among group members. The possibility of this last "group level" factor having an influence on cue utilization is the focus of the present study.

According to the common knowledge effect hypothesis, the group will use a different judgment policy if a cue is always shared than if it is always unshared. More specifically, the cue utilization ("weight") for a cue will increase as the number of group members who possess the cue before the group discussion increases.

The common knowledge effect hypothesis can be tested by comparing the utilization of each cue in the judgments of groups with different levels of sharing for that cue. According to the hypothesis, the standardized simple regression coefficient for a cue will be larger for a group in which all three group members always receive that item of information than for groups in which only one member ever receives the cue. Cue utilization for the condition in which two members receive that item of information should fall somewhere between the other two conditions. We expected that the judgment policy would be an additive linear relationship between the cue set and the judgments, with no interactions or higher order relationships in the model. With relatively simple judgments being made by nonexperts, in a small amount of time, it seems unlikely that the judges would rely on complex judgment strategies. The validity of this hypothesis about the judgment policies can be evaluated empirically by testing the goodness of fit of more complex models.

Method

Overview

College student subjects met in three-member groups to make judgments about the grades received by 32 stimulus students in an introductory psychology course. Each group member first received information about a stimulus student and made an individual judgment about that student. When all three group members had completed their individual judgments, the stimulus case information and the members' judgments were collected. The group members then discussed that stimulus case until they came to a consensus group judgment. That group judgment was collected and the information about the next stimulus case was distributed to the members. The above steps were repeated for each of the 32 stimulus cases (and for four repeated stimulus cases, which were used as a test of judgment reliability).

Subjects

University of Colorado students participated in the study as partial fulfillment of a research experience requirement of introductory psychology courses. Subjects were randomly assigned to three-member groups, composing three experimental conditions and one control condition. A total of 120 subjects participated as members of 40 groups, with 10 groups assigned to each condition. Seventy-seven of the subjects were male and 43 were female, but no effects of subject gender (individually or in the composition of the experimental groups) were observed in any statistical analysis, and so we will ignore this variable in the remainder of the present report.

Materials

The 32 stimulus cases were selected out of a pool of 504 students who had completed one of several introductory psychology courses during the previous year. Information about these students had been collected both from university records and from a voluntary questionnaire that was distributed at the end of the semester.

Six items of information were chosen as cues for the judgment task. These items consisted of high school grade point average, Scholastic Achievement Test or American College Test percentile score (averaged across the mathematics and verbal sections), self-reported percentage of lectures and recitations attended for the course, self-rated enjoyment of the class, self-rated academic anxiety, and self-rated workload in other courses. These cues were chosen to represent both ability and motivational variables and different strengths of relationship with the received course grades (ecological validity).

The 32 judgment cases were chosen to maximize the variability of the sets of cues to make more easily discernible the relationship between each cue and the judgments that are made. Three individual case profiles were created for each target within each condition. The individual profiles consisted of a piece of paper containing labels and values for each of the included cues. The order of the listed cues varied between stimulus cases and between the three profiles for each stimulus case.

Within each condition, the six informational cues were distributed such that the same two cues were always shared by all three group members, two other cues were always shared by two of the group members, and the remaining two cues were always given to only one group member, with the specific group members receiving a particular unshared cue varying across stimulus cases. Thus, each group member always received four items of information about each stimulus case. If the group members pooled their information, all six cues would be available to them. Between conditions, the cues were distributed such that in the first condition, a given cue (e.g., high school grade point average) was always shared by all three group members, in another condition that cue was always shared by two group members, and in the third condition, that cue was always given to a single group member. Thus, between conditions, each cue varied in the level of sharing.

Finally, a "fully shared information" control condition was run, in which all three group members always received all six of the judgment cues. This control condition was not strictly equivalent to the three experimental conditions, as members in the control condition groups received more information before each judgment than did members in the experimental condition groups. However, one goal of the present research was to compare information utilization of groups with that of their constituent members. Only by presenting all six cues to a member before each of the 32 judgments could we obtain unbiased estimates of the relationship between each cue and that member's judgments, controlling for the other five cues.

Table 1 shows the distribution of information in each of the four conditions.

Procedure

The procedures were the same for all conditions, except for the differences in information distribution that were described above. Participants met in groups of three. Preliminary instructions stated that the experiment was designed to study committee decision making. Subjects were told that they would be making judgments about the grades received by actual introductory psychology students like themselves and that they would be basing their judgments on information about these students that was collected from university records and from a questionnaire.

The subjects were next asked to fill out the actual questionnaire from which the self-reported cues (enjoyment of the course, course attendance, and academic anxiety) were collected. They were told that they were completing the questionnaire to see for themselves how the information was collected.

Next, the subjects received a sheet describing each of the cues. They were also informed that the judgment cases would include a representative range of received course grades and that, as in a real committee, they might not receive all of the same information as the other group members.

Subjects were next asked to express their subjective judgment policies for the six included cues by assigning relative weights to the cues, summing to 100, such that the weight for each cue represented how important that cue would be in judging what grade a student received in the introductory psychology course.

The group then completed an example judgment. Each group member was given an individual judgment case containing example cue values. For this practice case, each group member received the same five cues. They were informed that they were to make as accurate an individual grade judgment as possible from the information given by circling one of the 12 grades listed on the bottom of the sheet (from F to A, including pluses and minuses, but excluding F−, F+, and A+). After all three group members had completed the example individual judgment and returned the individual judgment sheets to the experimenter, they were informed that they were to discuss the case until they had reached a group consensus about the grade received by that student in the course. They were told that after a consensus had been reached, one of the members, who had been arbitrarily chosen as the recorder, was to record the group judgment by circling one of the grades on a group judgment sheet. They were informed that they would be taped during the group discussion so that a record could be kept of the content of their discussion.

The subjects then completed the individual and group judgment phases for each of the 32 cases. The cases were presented to each group in a different random order. An audio recording was made of the group discussion portion for each case. After the 32 cases had been completed, four of the original cases were repeated as a check of judgment consistency. The subjects were not informed that they had judged these cases before. The groups completed the 36 judgments in a mean time of 54 min. There were no reliable differences between the mean completion times of the groups in the experimental and control conditions.

After all of the 36 individual and group judgments were completed, the subjects completed a short final questionnaire. In this questionnaire, they were asked to report how accurate they thought the group judgments were, to rank the group members in order of accuracy of member judgments, and to rank the group members in order of influence on the group judgments. In addition, subjects repeated the subjective judgment policy weighting measure. After they completed the final questionnaire, subjects were informed of the true purpose of the study and excused.

Table 1
Distribution of Information by Experimental Condition: Number of Group Members Sharing Each Cue

	Experimental condition			
Cue	1	2	3	Control
ACT or SAT percentile	1	3	2	3
Attendance percentage	2	3	1	3
High school grade point average	3	1	2	3
Enjoyment	2	1	3	3
Other workload	3	2	1	3
Self-rated anxiety	1	2	3	3

Note. ACT = American College Test; SAT = Scholastic Achievement Test.

Results

Overview

Our analysis is presented in an individual to group order. First, we apply regression analysis to capture some of the abstract characteristics of the judgment task and the judgment "policies" followed by members making individual prediscussion judgments. Then we turn to the analysis of the group judgment process and evaluate the extent to which the common knowledge effect hypothesis appears to be manifested in our results. Finally, we attempt to separate the impacts of the members' prediscussion judgments and the discussion contents on the group judgments, again relying on regression analysis to organize our analysis.

Member Judgments

Reliability. For the member judgments, the average correlation between the first and second judgments for the four stimulus cases that were repeated was .57 for members in the experimental groups and .74 for members in the control group, both of which were reliably greater than zero, $F(1, 29) = 33.0$, $SD = 0.65$, $p < .0001$, and $F(1, 9) = 32.7$, $SD = 0.52$, $p < .001$, respectively.

There was no difference in average member judgment reliability between experimental conditions, $F(2, 27) = 1.28$, *ns*, or between members in the experimental groups and members in the control groups, $F(1, 37) = 1.83$, *ns*.

Member judgment policies. If group members were utilizing cue information in a linear additive manner, linear regression can be used to ascertain the judgment policies of the individual group members. Regressing a group member's 32 grade judgments on the values of the corresponding stimulus cues provides a model of cue utilization. Such a procedure cannot be performed for the individual subjects in the three experimental conditions, because those subjects were missing information for every judgment; each subject only received four of the six cues for each case. In their judgments, subjects could not utilize information that they did not have and the regression estimates of judgment policy would be confounded with the experimental manipulation.

However, member judgment policies could be obtained for the subjects in the control condition, in which all three group members always received all six of the stimulus cues for every stimulus case. The absolute average standardized regression coefficients for these group members are shown in Figure 1. The analysis shows that the grade judgments of individual subjects were influenced most heavily by the high school grade point average and aptitude test percentile cues, somewhat less by attendance and enjoyment, and very little by other workload and self-rated academic anxiety.

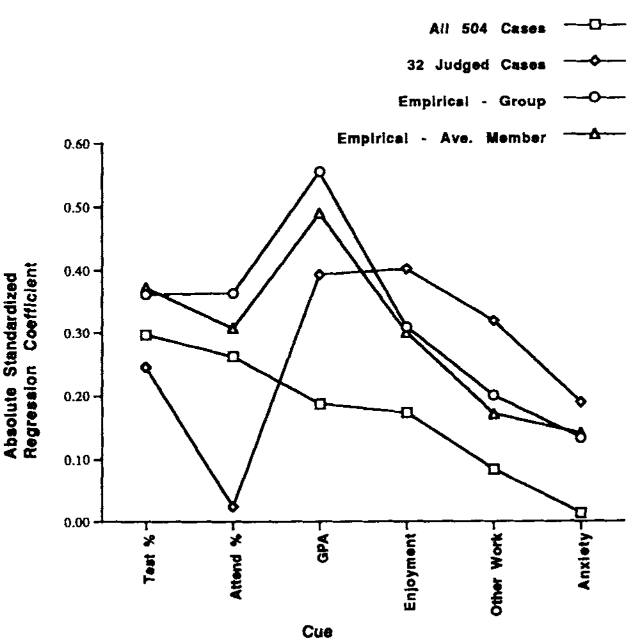

Figure 1. Best linear models versus mean group and member empirical judgment policies (control condition groups). (Ave. = average; GPA = grade point average.)

Throughout the analyses, we relied on average results to summarize individual and group judgment policies. In every case, we conducted fine-grained analyses of policies at the appropriate level and observed some variation in the patterns of weights obtained. For example, cluster analyses of the individual judgment policy weights obtained from the control group subjects (who received full six-cue information sets on each judgment trial) revealed three groupings of subjects according to their policies. However, the averaged data provides a fair representation of the typical and predominant patterns of cue utilization in judgment in every case in which we report statistics at that level.

One question of interest is how well the judgment policies of the individual group members matched the optimal linear model weights, summarizing the true relationships between variables. Did the group members use the available information correctly? This question was answered by comparing the individual judgment policies with the best model for the data from which the judgment stimulus cases were chosen. Five hundred four cases in the data set had values for all six of the informational cues that were presented to the subjects. The optimal linear weights were obtained by regressing the actual grades received by the students on the six cues: High school grade point average, aptitude test percentile, attendance, self-rated enjoyment of the course, self-rated academic anxiety, and self-rated other workload. The coefficients in this model represented the ecological validity of the cues—how the cues truly were related to the course grades. The absolute standardized regression coefficients for the 504 cases are also shown in Figure 1. The parameter estimate for other workload was negative, whereas the parameter estimate for anxiety was not reliably different from zero, $F(1, 497) = 0.10$, *ns*. The model including the six cues as predictors accounted for 29% of the variance in introductory psychology course grades, $F(6, 497) = 33.1$, $p < .0001$.

We also tested for more complex relationships between the cues and the criterion. A test of all of the two-way interactions between pairs of cues revealed reliable positive interactions between the high school grade point average and aptitude test percentile cues, $F(1, 482) = 7.92$, partial $r^2 = .02$, $p < .01$, and between the other workload and the anxiety cues $F(1, 482) = 5.60$, partial $r^2 = .01$, $p < .05$. In another model testing for the quadratic effect of each cue, we found a positive quadratic effect of high school grade point average, $F(1, 491) = 10.1$, partial $r^2 = .02$, $p < .01$. However, adding the two interactions and the quadratic effect of high school grade point average to the model predicting course grades from the six cues only reduced the error by an additional 4.4%. Therefore, we concluded that a simple linear model adequately describes the true relationships in the data.

The best linear model for the 32 to-be-judged cases is also graphed in Figure 1. The plotted points show the absolute standardized regression coefficients from a model regressing introductory psychology course grade on all six cues. As was noted previously, the 32 cases were chosen to maximize variability in the values of all six cues. The resulting cue-to-grade relationships were somewhat different from the optimal model for all

504 students. The best linear model for the 32 cases accounted for 42% of the variance in course grades, $F(6, 25) = 3.00$, $p < .05$. The coefficients for aptitude percentile, attendance percentage, and other workload were not significantly different from zero. The coefficients for other workload and anxiety were negative. For these 32 cases, the correlation between the predictions from the best linear model for all 504 students and the predictions from the best linear model for the 32 to-be-judged students was $r = .79$, $p < .0001$.

Again, the average individual judgment policies of the control group subjects can be compared with these optimal linear weights. As shown in Figure 1, the subjects relied more on each of the cues than they should have, according to the best linear model. Such an overutilization makes sense, as the judges only had these six cues on which to base their judgment, whereas in the real world there likely are many other variables that covary with introductory psychology grades. Although the overutilization of cues leads to differences between the judged grades and the actual grades, the correlation between judged grades and actual grades would still be high if the two patterns of cue usage are similar.

A comparison of lines in Figure 1 shows that the judgment policies of the individual subjects followed fairly closely the form of the optimal linear weights. The exception to this generalization, the exceptionally high weights for the high school grade point average cue, probably represents a cue–response scale commensurability effect of the type that has been frequently observed in other judgment studies: When a cue is presented on the same scale as the response, it tends to receive a large weight in the judgment policy (Slovic, Griffin, & Tversky, 1990; Slovic & MacPhillamy, 1974). Otherwise, the judgment policies of the individual members in the control groups followed a pattern similar to the best linear model. The average correlation between the 32 judgments of each control group member and the grade for each of the 32 cases that was predicted by the best linear model was .68, which was reliably different from zero, $F(1, 9) = 287.6$, $SD = 0.15$, $p < .0001$. It appears, therefore, that these subjects were using the available information in a (linear, additive) way that reflects the true relationships between the variables.

Insight. We were also interested in whether the subjects were aware of how they were using the provided information. All of the subjects were asked to give their subjective judgment policies both before and after they made the grade judgments. In each case, the subjects were asked to divide 100 points between the six stimulus cues according to how important they thought each cue should be in the grade judgment. The means for the postjudgment subjective judgment policies are shown in Figure 2. A comparison of these mean subjective influence ratings with the average member judgment weights suggests that subjects are unaware of the relatively high influence of the high school grade point average and aptitude test percentile cues on their judgments.

Accuracy. As a measure of individual judgment accuracy, a correlation was calculated for each subject between his or her individual grade judgments and the actual grades across the 32 stimulus cases. All analyses involving these correlations, and all others in the study using correlations as the dependent variable, were performed after first converting the correlations to Fisher's

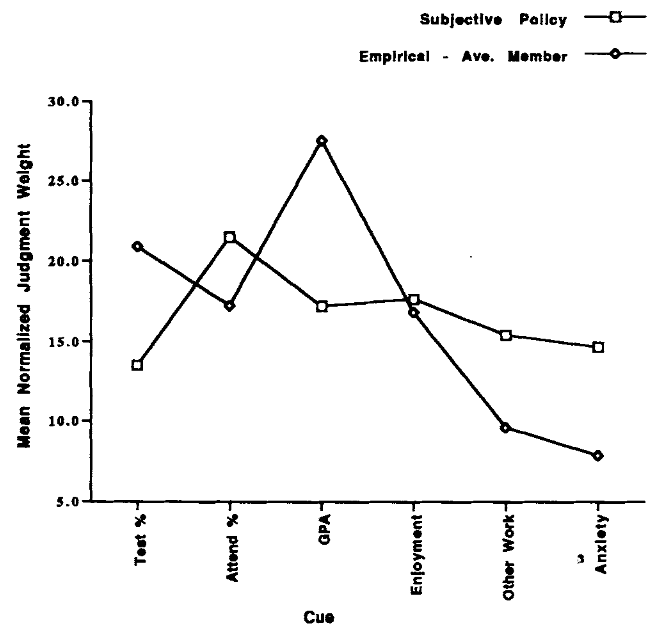

Figure 2. Mean postexperiment subjective judgment policy (control condition groups) versus mean member empirical judgment policy (normalized to sum to 100). (Ave. = average; GPA = grade point average.)

z scores, to correct for heterogeneity of error variance (Judd & McClelland, 1989). The correlations also were averaged within each group, because the levels of accuracy for the members of a particular group are likely not to be independent from each other.

The average correlations between individual judgments and actual grades for each condition are presented in Table 2. All of the average correlations are reliably greater than zero. The accuracy of the group members' judgments is reliably different among the three experimental conditions, $F(2, 27) = 12.3$, $R^2 = .48$, $p < .001$. This between-conditions difference in member judgment accuracy was expected, because group members in the different conditions received information of different cue validity, on average, because of the manipulation of information sharing. However, it is surprising that the subjects in Condition 2 were the least accurate; those subjects received aptitude percentiles and attendance percentages for every stimulus case, the two cues that should have been the most useful in making judg-

Table 2
Mean Member Judgment Accuracy by Condition

	Experimental condition			
Accuracy measure	1	2	3	Control
Correlation: Member judgment vs. actual grade	.41	.23	.32	.38
Correlation: Member judgment vs. grade predicted by best linear model	.59	.60	.56	.68

ments. The accuracy of group members in the experimental conditions was not significantly different from the accuracy of group members in the control condition, $F(1, 38) = 2.16$, *ns*. Even though the control group members received more information about each stimulus case (six cues each), they were not more accurate than the experimental group subjects (who received four cues each).

An alternative measure of accuracy considers the magnitude of the difference between each judgment and the corresponding criterion. Judgments that had patterns of relative differences similar to those of the criterion grades, and therefore a high accuracy correlation, could still be inaccurate in terms of overestimating or underestimating the criterion grades. To explore the absolute accuracy of member judgments, we averaged the absolute differences between each member's judgments and the corresponding criterion grades. This absolute difference measure of accuracy showed the same pattern of differences between conditions, with the groups in Condition 2 being the least accurate and with no mean difference between groups in the experimental conditions and groups in the control conditions.

The social judgment theory analysis suggests another conceptualization of judgment accuracy. According to social judgment theory, an accurate judge would use a judgment policy that is similar to the validity of cues in the environment. Thus, a measure of judge accuracy over multiple judgments would compare the predictions of that judge's judgment policy to the predictions of the best model of cue validity. However, we were not able to estimate the judgment policies of individuals in the experimental conditions. Instead, we computed correlations between the predictions of the best linear model for all 504 cases in the data set and the 32 judgments of each group member. These correlations measure how similar the judgments of each member are to the judgments that he or she would make if the linear relationship of cues to the criteria were known. Table 2 shows the means of these correlations by condition. The mean correlation between best model predictions and member judgments did not differ among experimental conditions or between the experimental conditions and the control condition.

Group Judgments

Reliability. The average correlation between the group judgment for the first and second repetition of the four repeated stimulus cases was .87 for the experimental groups and .84 for the control groups. Both average reliabilities were significantly greater than zero, $F(1, 29) = 65.6$, $SD = 0.79$, $p < .0001$, and $F(1, 9) = 40.0$, $SD = 0.58$, $p < .001$, respectively. There were no differences in average group reliability between groups in the 3 experimental conditions, $F(2, 27) = 0.66$, *ns*, or between the experimental groups and the control groups, $F(1, 37) = 0.03$, *ns*. For the experimental groups, the group judgments were more reliable, on average, than were the judgments of their members, $F(1, 29) = 17.5$, $SD = 0.68$, $p < .001$; this difference was not significant for the control groups, $F(1, 9) = 2.14$, *ns*.

Group judgment policies (common knowledge effect). At the level of group judgments, regressing the group grade judgments on the six stimulus cue values provides a measure of the influence of each cue on the group judgment. The judgment policies of the full sharing control groups again provide a measure of information usage that is not confounded by differences in information distribution. The mean standardized regression weights for these groups are presented in Figure 1. For these groups, the group judgment policies are very similar to the member judgment policies.

The common knowledge effect hypothesis implies that the experimental groups' judgment policies should differ from condition to condition because of differential distribution of information. Recall our prediction is that the more members who are given a cue, the greater the impact on the group judgment (indexed by the standardized regression weight). To test this hypothesis, a one-way analysis of variance (ANOVA) was performed for each of the six cues. In these analyses, the absolute standardized regression weight for each stimulus cue was used as the dependent measure, with number of group members holding that cue as the between-groups factor. (It should be noted that these six ANOVAs are not independent tests of the hypothesis, as the influence given to any one cue by a group likely affects the influences of the other five cues.)

The relationships between cue distribution and mean regression coefficients for each cue are shown in Figure 3. Significant linear effects of cue "sharedness" on the weight of that cue in the group judgment policy were found for 4 of the 6 stimulus cues. The effect was found for aptitude test percentile, $F(1, 27) = 26.7$, partial $r^2 = .50$, $p < .0001$, attendance percentage, $F(1, 27) = 58.6$, partial $r^2 = .68$, $p < .0001$, high school grade point average, $F(1, 27) = 44.9$, partial $r^2 = .62$, $p < .0001$, and enjoyment, $F(1, 27) = 23.3$, partial $r^2 = .46$, $p < .0001$. For these cues, more sharing led to greater impact.

If there is a direct linear relationship between level of sharing and influence of a cue, then the coefficient when that cue was shared by two group members should fall near the mean of the

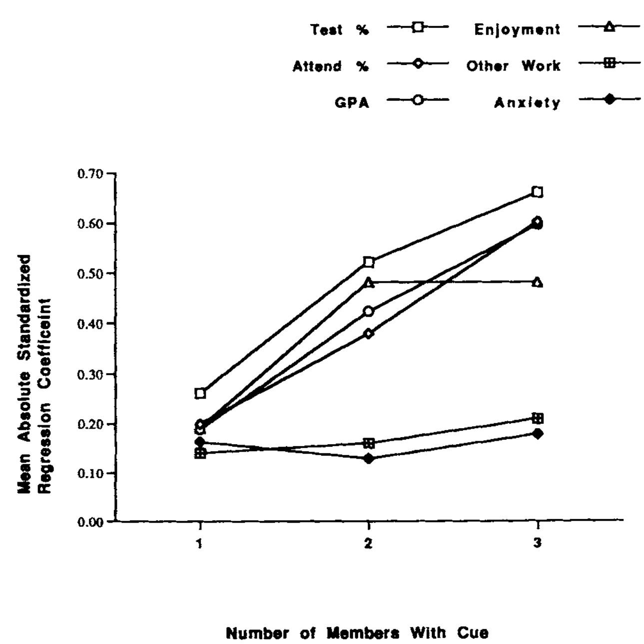

Figure 3. Effect of cue sharing on group judgment policy weights. (GPA = grade point average.)

other two levels of sharing. Enjoyment was the only cue for which the influence was reliably different when shared by two group members than the average of the other two levels of sharing, $F(1, 27) = 7.81$, partial $r^2 = .22$, $p < .01$. There were no reliable differences between levels of sharing for the influence of other workload or self-rated anxiety cues.

Accuracy. To assess group judgment accuracy, the grade judgments of each group were correlated with the actual stimulus case grades. The average correlations for each condition are reported in Table 3. Again, there were reliable differences in accuracy due to experimental condition, $F(2, 27) = 10.97$, $R^2 = .72$, $p < .001$. The average group accuracy correlations of the experimental conditions and the control condition were not significantly different, $F(1, 38) = 0.39$, *ns.*

We also looked at the average absolute group judgment accuracies, computed as the average absolute difference between a group's judgments and the criterion grades. Again, the pattern of between-groups differences for these scores was similar to that of the judgment–criterion correlations.

Table 3 also shows the mean correlations between the group judgments and the grades predicted by the best linear model. These correlations give an indication of how well the groups were using the case information, in comparison with the true relationships between cues and criterion grades. No significant between-conditions differences were found.

Did the judgments improve as a result of group discussion? To address this question, we compared the group accuracy correlation of each group with the average accuracy correlation of the group's three members. The difference between the accuracy of the group judgments and the accuracy of the individual member judgments was reliable for the experimental conditions, $F(1, 29) = 44.8$, $SD = 0.08$, $p < .0001$, and marginally reliable for the control groups, $F(1, 9) = 4.25$, $SD = 0.09$, $p = .07$. On average, the judgments of the groups were more accurate than the judgments of their constituent members. This difference is easily explained as due to the cancellation of random judgment error through aggregation (cf. Davis, 1969). Simply by statistically pooling their judgments, by averaging them, for example, the three group members would be likely to improve their judgment accuracy.

Of more interest, then, is whether the group judgments were more accurate than the composite judgment obtained by averaging the three member judgments for each case. That is, did the group discussion add any more to judgment accuracy than would have occurred had the members' judgments about each case simply been averaged? The average correlation between the mean of the three member judgments and the actual grades is shown in Table 3. The accuracy of the group judgments was not reliably different from the accuracy of the average of the member judgments for either the groups in the three experimental conditions, $F(1, 29) = 0.90$, *ns,* or the groups in the control condition, $F(1, 9) = 0.01$, *ns.* The group judgments were not more accurate than a statistical pooling of the three members' judgments.

Table 3
Mean Group Judgment Accuracy by Condition

Accuracy measure	Experimental condition 1	2	3	Control
Correlation: Group judgment vs. actual grade	.50	.29	.41	.43
Correlation: Mean member judgment vs. actual grade	.53	.29	.41	.44
Correlation: Group judgment vs. grade predicted by best linear model	.72	.75	.72	.73
Correlation: Mean member judgment vs. grade predicted by best linear model	.76	.73	.73	.75

To gain a clearer understanding of the relationship between group and member judgment accuracy, we ranked each group's judgment accuracy correlation in relation to the judgment accuracy correlations of its constituent members. A ranking of 1, for example, indicated that the group judged more accurately than any of its members, a ranking of 2 indicated that the group was less accurate than one member, and so on. The mean group versus member accuracy rank was $M = 1.83$ for the experimental groups and $M = 1.90$ for the control groups. In the experimental conditions, 10 of 30 groups were more accurate than any of their members, whereas none was less accurate than all of its members. In the control groups, 4 of 10 groups were more accurate than all of their members and 1 was less accurate than all of its members.

Sniezek and Henry (1989) found that groups with initial judgment disagreement between members made more accurate group judgments than did groups whose members were in agreement. In the present study, groups may have pooled more information, spent more time discussing the case, or paid more attention to unshared information when their initial member judgments were in disagreement. We measured member disagreement about each case by taking the mean of the three absolute differences between pairs of member judgments. We then correlated this measure with the absolute difference between the group judgment and the criterion grade across the 32 cases. These correlations measured the relationship between average member judgment disagreement and group judgment accuracy for each group. Contrary to Sniezek and Henry's findings, the mean disagreement–accuracy correlation was not significantly different from zero for groups in any of the experimental or control conditions. However, the present judgment paradigm differed from the one used by Sniezek and Henry. Their groups made estimates of unfamiliar quantities, the frequency of deaths from various causes. Therefore, a group member whose judgment was discrepant from the other members may have been more knowledgeable and therefore more accurate. A member who was knowledgeable probably would have little trouble convincing the other group members of his or her opinion. In addition, Sniezek and Henry used the geometric mean as a measure of judgment accuracy, which may have had properties different from our own accuracy measures.

Relationship Between Individual and Group Judgments

Another issue of interest is whether the stimulus cues have any effect on the group judgments, beyond the effect that they have on the individual judgments. Do the groups simply pool the individual member judgments, or do the values of the cues have an additional effect, even when controlling for the individual judgments? This question was addressed by regressing the group judgments on the individual judgments of each of the three group members, along with the set of six stimulus cues. For the experimental conditions, only the coefficient for attendance was reliably greater than zero, $F(1, 29) = 8.3$, $SD = 0.12$, $p < .01$. In general, the information that was provided to the groups did not have an effect on the group judgments above and beyond its mediated effect through the group members' judgments.

Even though they are not greater than zero, on average, the direct influences of the cues could be tested to determine whether they show an effect of level of sharing. Again, mean absolute values of the standardized regression coefficients were used as a measure of cue influence. The mean regression weights for each condition are shown in Figure 4. The test of the difference between the unshared (held by one member) and fully shared (held by three members) conditions was not reliable for any of the cues. In this judgment task, information distribution does not appear to have an effect on the group judgment policies, above and beyond its effect mediated through the individual judgments.

How, then, are the group members combining their judgments into a group judgment? They might simply be averaging their member judgments. The mean correlation between each group's judgments and the average of its members' judgments was .94 for the groups in the experimental conditions and .96 for the groups in the control conditions, both reliably different from zero. Although the case is strong for an averaging combination rule, some of the groups seemed to be using a different strategy on at least some of the cases. From their group discussion, they seemed to be averaging the closest two members' judgments and ignoring the third member's judgment. The mean correlation between the groups' judgments and the mean of the two closest judgments of their constituent members was .87 for the groups in the experimental conditions and .93 for the groups in the control condition. The difference in fit between the averaging combination rule and the "closest two" combination rule was reliable for both the groups in the experimental conditions and the groups in the control condition, $F(1, 29) = 292.8$, $SD = 0.25$, $p < .0001$, and $F(1, 9) = 12.7$, $SD = 0.23$, $p < .01$, respectively. We also tested a median combination rule. Groups might simply be using the middle member judgment as their group judgment. The mean correlation between the median judgment and the group judgment was $r = .90$ for the groups in the experimental conditions and $r = .93$ for the groups in the control condition. The difference between the fit of the averaging combination rule and the fit of the median combination rule was reliable for the experimental conditions, $F(1, 29) = 14.9$, $SD = 0.18$, $p < .001$, and marginally reliable for the control condition, $F(1, 9) = 4.67$, $SD = 0.14$, $p = .059$. Thus, the general combination rule for these groups seems to be best described as a simple average of the members' judgments.

We can also examine the present data to determine whether there are any judgment shifts from the initial individual judgments to the group judgment across the 32 cases (Brown, 1986; Myers & Lamm, 1976). There is a slight and consistent tendency for the slope of the regression line to predict the group judgment from the average of the three member judgments to be greater than unity (average slope for the four experimental conditions of 1.05). The interpretation of this small shift is ambiguous as it is both a group polarization shift and a shift toward increased accuracy.

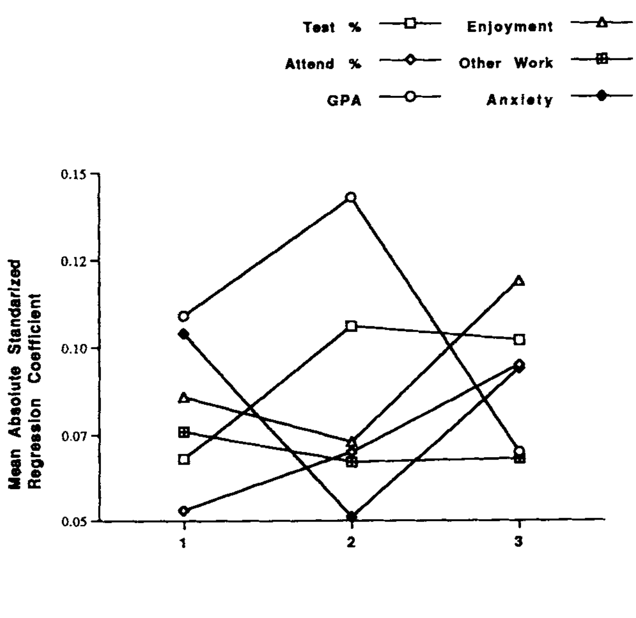

Figure 4. Effect of cue sharing on group judgment policy weights, controlling for member judgments. (GPA = grade point average.)

Contents of Group Discussion

The following transcript excerpts provide two examples of typical group discussions, the first by a group in experimental Condition 2 and the second by a group in the control condition.

Example Group Discussion 1 (Condition 2).

C plus.

I gave him a D.

I gave him a B minus.

What'd you know that we didn't?

He put down the ultimate lowest for enjoyment of the class.

Oh, really?

So, I'm thinking he must have gotten a bad grade. When he didn't do as well as he'd like, he put down the lowest.

Yeah, he gave a zero, huh?

Yeah.

What'd you. . . . Did you have his G.P.A. or anything?

Three point one, and he went to class 82 percent of the time, or something like that.

Yeah, 82. . . . And if you go to class 82 percent of the time. . . .

I just. . . . That zero just kind of freaks me out. Why would someone hate a class that much?

He had a five for anxiety.

Five for anxiety? Which is average, right?

Yeah.

So, I don't know. I think he must have gotten a bad grade. I was almost going to give him an F, because he put down a zero.

D plus. Give him a D plus.

I gave him a C or something.

C minus?

How about that?

Yeah.

All right.

Example Group Discussion 2 (control condition).

I gave him a B plus, because his anxiety was pretty high, and his enjoyment was high, but his workload wasn't that bad. And he seemed like a pretty good student.

I gave him a B. It just seemed like he was a good student, and he enjoyed it. His attendance wasn't the best, but I just thought he deserved a B, got a B in the class. So. . .

B plus?

Yeah.

A simple content analysis was performed on the group discussions of 36 of the groups (4 tapes were unusable). Tapes were not coded for 1 group in Condition 1, 1 group in Condition 2, and 2 groups in the control condition. Two coders analyzed the discussions. Each coder listened to half of the discussions within each condition. The coder noted each time that one of the six cues was mentioned by a group member during discussion of each of the 32 cases. As a rough test of the reliability of the coding scheme, both coders coded discussions of 3 of the cases by 8 different groups (2 groups from each condition). During discussion of those 24 cases, one or the other coder noted 92 different cue mentions. The coders were judged to disagree on a coded event any time that one coder marked more mentions of a particular cue than the other did. The coders were in exact agreement on 93.5% (86) of the coded events. Therefore, we are confident that the coding scheme was sufficiently reliable.

Group members almost always pooled their own judgments. On average, 2.93 (of the 3) members' judgments were mentioned during the discussion of each case. The pooling of member judgments did not differ significantly between experimental conditions or between the experimental groups and the control groups.

The pooling rates that are presented below slightly underestimate how much pooling actually occurred during the discussion of a typical case. When the initial judgments of all three group members were in agreement, the members rarely pooled any cues at all. On average, all three group members agreed on 2.71 of the 32 judgments.

The primary variable of interest was whether a particular cue was pooled at least once during the discussion of a case. After a cue had been pooled once, that information was available to all three group members. Any more pooling of that cue would therefore present redundant information. The groups in the experimental conditions discussed a mean of 3.87 cues (of six) for each case. The groups in the control condition discussed a mean of 2.97 cues per case. The difference between the experimental and control groups was reliable, $F(1, 34) = 9.25$, $p < .005$. Groups in the three experimental conditions did not differ in their pooling rates, $F(2, 25) = 1.23$, *ns*.

Repeated pooling of a cue might have an effect. Group members might consider a cue that is pooled multiple times during discussion of a single case to be more important than a cue that is only pooled once. Therefore, we also considered the total number of times that each cue was pooled during the discussion of each case. Including repeated pooling of the same cue, groups in the experimental conditions discussed a mean of 5.31 cues for each case. Groups in the control condition discussed a total of 4.05 cues, on average. The difference in total pooling between the experimental groups and the control groups was only marginally significant, $F(1, 34) = 3.27$, $p = .08$. The total pooling rates of the three experimental conditions did not differ, $F(2, 25) = 0.52$, *ns*.

In most cases, analyses involving the two variables revealed a similar pattern of results. In the interest of economy, therefore, we only report the analyses involving total pooling rates when they differ qualitatively from those involving the categorical pooling variable. In the between-groups analyses that follow, pooling was averaged across the 32 prejudgment discussions. Thus, the pooling rate for a cue (for the categorical coding) was expressed in terms of the proportion of those 32 discussions during which that cue was mentioned.

First, we looked for a common knowledge effect for information pooling. The mean pooling rates by condition are shown in

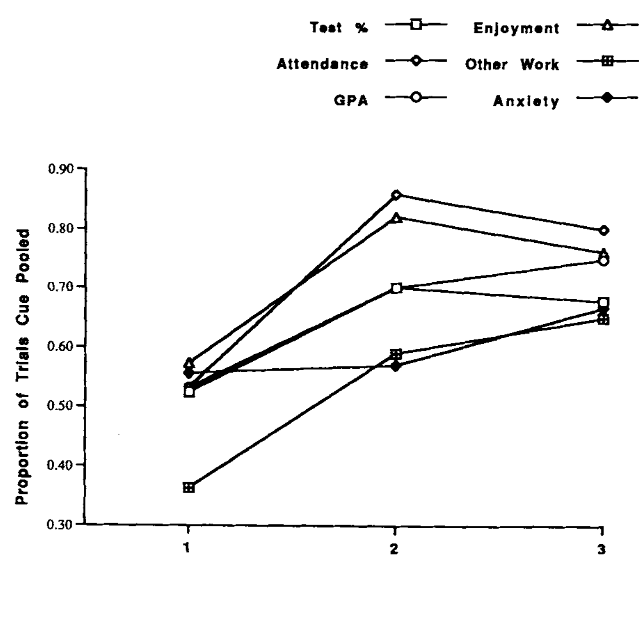

Figure 5. Effect of cue sharing on rate of cue pooling. (GPA = grade point average.)

Figure 5. Cue distribution among members and the discussion of an individual cue were related for five of the six cues. The linear effect of sharing was marginally reliable for the aptitude test percentile cue, $F(1, 25) = 4.11$, partial $r^2 = .14$, $p = .053$), and reliable for attendance percentage, $F(1, 25) = 14.9$, partial $r^2 = .37$, $p < .001$, high school grade point average, $F(1, 25) = 5.97$, partial $r^2 = .19$, $p < .05$, enjoyment, $F(1, 25) = 6.94$, partial $r^2 = .16$, $p < .05$, and other workload cues, $F(1, 25) = 27.8$, partial $r^2 = .53$, $p < .0001$. In addition, the quadratic test comparing the shared-by-two condition with the average of the unshared and fully shared conditions was reliable for attendance percentage, $F(1, 25) = 10.0$, partial $r^2 = .29$, $p < .001$, and enjoyment, $F(1, 25) = 6.05$, partial $r^2 = .19$, $p < .05$, and marginally reliable for other workload, $F(1, 25) = 3.10$, partial $r^2 = .11$, $p = .091$. All of the cues besides anxiety were discussed more when they were shared by all three group members than when they were unshared. However, attendance, enjoyment, and other workload were discussed the most when they were shared by two of the three group members. The pattern is only slightly different for the total pools per judgment measure. With that dependent measure, the quadratic effect of cue distribution is not reliable for the other workload cue, $F(1, 25) = 1.08$, *ns*.

The pooling rates suggest that the effect of cue distribution on cue impact could be due to increased pooling of shared cues. Alternatively, group members might be more likely to pool the cues that have an impact on the group judgments. To explore these possibilities, we tested the effect of each cue's distribution on its impact on the group judgments (mean standardized regression weight), controlling for the cue's pooling rate. In this test, the linear effect of sharing was reliable for test percentile, $F(1, 24) = 13.8$, partial $r^2 = .37$, $p < .005$, attendance percentage, $F(1, 24) = 25.9$, partial $r^2 = .52$, $p < .0001$, high school grade point average, $F(1, 24) = 23.3$, partial $r^2 = .49$, $p < .0001$, and enjoyment, $F(1, 24) = 10.5$, partial $r^2 = .30$, $p < .005$. The quadratic effect of distribution on cue impact, controlling for pooling, was not significant for any of the cues. Thus, for all 4 cues that originally showed a common knowledge effect (impact of the group judgment), that effect remained when we controlled for between-conditions differences in pooling. However, the quadratic effect of distribution for the enjoyment cue disappeared. Thus, the high impact of enjoyment when it was shared by two group members may have been related to the frequency of pooling of enjoyment in those groups.

Within these same analyses, the test for the effect of pooling on cue impact was also of interest. Each of these tests would show whether the frequency of pooling of a cue was related to the impact of that cue, within levels of sharing. The test was positive and reliable for test percentile, $F(1, 24) = 6.68$, partial $r^2 = .22$, $p < .05$, attendance percentage, $F(1, 24) = 9.30$, partial $r^2 = .28$, $p < .0001$, and enjoyment, $F(1, 24) = 10.5$, partial $r^2 = .27$, $p < .01$, and marginally reliable for high school grade point average, $F(1, 24) = 3.91$, partial $r^2 = .14$, $p = .060$. Within levels of sharing, a cue had more impact on a group's judgments when it was pooled more frequently. For the test percentile, $F(1, 24) = 0.33$, *ns*, and attendance percentage, $F(1, 24) = 2.37$, *ns*, cues, however, the effect of mean total pooling was not reliable, suggesting that for those cues, what mattered was whether the cue was pooled during each discussion, not how many times it was pooled.

We also tested for interactions between cue distribution and pooling. These interaction terms test whether the common knowledge effect was weaker (or stronger) for groups that pooled a cue more frequently. None of the interactions was reliable, for either the pooling frequency or the total pooling measures. Thus, the effect of distribution on cue impact does not appear to vary depending on amount of pooling.

We can also ask whether discussion of a cue changes its impact on a group's judgments. If a cue was pooled during discussion, did that information have more of an impact on the subsequent group judgment? Within each group, we estimated regression coefficients for six interaction terms. Each of these terms represented the interaction between the value of a cue and the pooling variable for that cue. Thus, each of these interactions showed whether the impact of a cue depended on whether it was pooled. We then tested whether the means of the six interaction term regression coefficient estimates were significantly different from zero. None of these tests were reliable for either the experimental condition groups or the control condition groups. On average, then, cue impact did not depend on pooling. Only for anxiety did the Cue × Pooling interaction vary significantly between groups in the different experimental conditions, $F(2, 17) = 7.93$, $R^2 = .48$, $p < .005$. Anxiety had the most (negative) impact on group judgments when it was pooled during discussion by groups in the condition in which that cue was fully shared. In general, however, pooling did not influence cue impact, whatever the cue distribution among group members.

We were also interested in changes in pooling rates over trials. Did the group members learn which of the cues were unshared and therefore begin to pool those cues more often, relative to the other cues? Within each group, pooling for each cue was correlated with trial number. Each of these correlations showed whether a cue was pooled more or less often as the group worked through the 32 judgments. These correlations were then averaged across groups. On average, the cues tended to be pooled less often in later trials. For groups in the experimental conditions, the average correlations were significantly less than zero for other workload (mean $r = -.14$), $F(1, 27) = 17.2$, $p < .001$, and anxiety (mean $r = -.11$), $F(1, 27) = 6.62$, $p < .05$, and nearly so for attendance percentage (mean $r = -.06$), $F(1, 26) = 3.96$, $p = .057$. For groups in the control condition, the average correlations were negative and significant for aptitude test percentile (mean $r = -.26$), $F(1, 7) = 5.74$, $p < .05$, attendance percentage (mean $r = -.32$), $F(1, 7) = 10.6$, $p < .05$, and enjoyment (mean $r = -.12$), $F(1, 6) = 10.0$, $p < .05$, and marginally significant for high school grade point average (mean $r = -.13$), $F(1, 7) = 4.05$, $p = .084$. Differences in the degrees of freedom are due to the lack of variation in pooling of some cues by some groups. None of the average correlations were significantly greater than zero.

More important, none of these average correlations showed an effect of cue distribution. Changes in pooling over time did not depend on the distribution of that cue across the group members. Thus, group members did not adjust their pooling based on knowledge about which cues were shared or unshared on every trial. However, for three of the six cues, the change in pooling over time did differ between groups in the experimental conditions and groups in the control condition. Pooling of test percentile, $F(1, 33) = 7.23$, $R^2 = .18$, $p < .05$, and attendance, $F(1, 33) = 11.4$, $R^2 = .26$, $p < .01$, decreased faster among groups in the control condition, whereas pooling of other workload, $F(1, 34) = 7.66$, $R^2 = .18$, $p < .01$, decreased faster for groups in the experimental conditions.

Discussion

The present results demonstrate the hypothesized common knowledge effect. In the context of a group judgment, three-member groups weighed the same information about target students more heavily when it was shared than when it was unshared. Even though groups in all three experimental conditions possessed exactly the same information about the target cases, their use of that information depended on its prediscussion distribution among the individual group members. The predicted effect was highly reliable for four of the six informational cues that were presented for the target cases: aptitude test percentile score, lecture attendance percentage, high school grade point average, and course enjoyment. Each of these four cues had reliably more influence when it was known by all three group members before group discussion than when it was known by only one group member. In addition, the other workload cue showed a slight linear trend, although it did not approach significance.

We want to emphasize the distinctive advantages of our seminal application of social judgment theory's multiple regression analysis framework. We are able to clearly measure, on an interpretable scale, the relative impact of various sources of information (cues) on individual and group judgment. We can analytically separate and compare the strength of mediation through individual-member judgments versus group discussion. And we can represent alternative theories about the group process either in comprehensive sets of equations or in path analysis diagrams. Although we have noted several limitations of the range of applicability of the method, under those conditions where it is appropriate, it performs like a surgeon's scalpel in comparison with the butter knife of conventional approaches to the analysis of the impact of information on group decisions.

Self-rated anxiety was the only informational cue that clearly was not used according to the hypothesis. The influence of the anxiety information did not depend on experimental condition. It could be that anxiety was simply not considered to be useful information in any condition. It did, on average, have the lowest influence of any of the six cues on both member and group judgments. In addition, anxiety was probably a difficult piece of information for groups to use, because it has no clear valence. It is not apparent whether higher self-reported anxiety about academics would be associated with higher or lower course grades or whether that relationship would be curvilinear. In fact, the influence of anxiety over all 32 cases was slightly positive for some groups and negative for others. Uncertainty about the nature of the relationship between anxiety and course grades within the group could have lowered the cue's influence to the point that the experimental manipulation had no effect. Interestingly, self-reported workload in other courses, the only cue with a clear negative valence, was the only other cue that did not clearly follow the expected pattern. Perhaps the common knowledge effect is strongest for information that has a clear, positive valence in the context of the judgments that the group is making.

One issue of interest concerned the linearity of the hypothesized common knowledge effect on those cues whose influence did reflect their distribution across members. That is, does each increase of one in the number of group members sharing a cue lead to an equal increase in the average influence of that cue on the group's judgments? For the course enjoyment cue, the effect of changing the prediscussion information distribution made a larger difference in average influence when moving from the condition in which enjoyment was known by one group member to the condition in which it was known by two than when moving from the condition in which it was known by two group members to the condition in which it was known by all three group members. The comparison was not reliable for any of the other cues. In general, the common knowledge effect is approximately linear, at least for three-member groups. Further study would be necessary to determine whether something about self-rated enjoyment causes that information to be used differently than the other cues, or whether the nonlinear effect is simply the result of something peculiar in the overall information distribution among the three experimental conditions.

Our regression analyses showed that the stronger influence of fully distributed information cannot easily be explained by differential information pooling. Stasser and his colleagues (Stasser, Taylor, & Hanna, 1989; Stasser & Titus, 1985, 1987) have found that distributed cues are more often mentioned by group members during the group discussion phase of a group judgment task. They have used this information pooling effect to explain their finding that inequitable information distribution can bias the outcome of a group judgment. Direct comparisons with Stasser's results are probably inappropriate because there are many differences between the nature of the task presented to subjects and the experimental procedures in the two research programs. We should note that the present procedure was likely to maximize overall information pooling and therefore to minimize differences in pooling between cues. In each of the 32 grade judgments, individual group members in the experimental conditions only received four of the total of six possible cues, in contrast to Stasser's experiments, in which a larger set of cues had to be remembered by the group members, who were making a single judgment. Moreover, subjects in the present study made a series of 32 judgments, using the same six cues, after having been informed about what kind of information they might expect to see.

The content analysis of the group discussions provided some insight into the relationship between pooling and cue impact. Congruent with Stasser's argument, all of the cues except anxiety were discussed more often when they were fully shared than when they were unshared. Surprisingly, however, 3 of those cues were pooled the most often when they were shared by only two of the three group members. This may indicate that subjects are cognizant of which cues are not fully shared and therefore need to be pooled.

Even though cue distribution and pooling were related, that relationship does not appear to explain the common knowledge effect of distribution on cue impact. When pooling was held constant, the common knowledge effect was still significant for the same 4 cues. Were groups in different conditions to pool one of those cues at the same rate, the impact of that cue on the group judgments would increase monotonically with the level of sharing. Within groups in the same condition, however, the more a cue was pooled, the more impact it had (again, for the aptitude percentile, attendance percentage, high school grade point average, and enjoyment cues). Obviously, the direction of causality in this relationship cannot be specified. It could be, as Stasser suggested, that pooling leads to increased cue impact. On the other hand, a cue that is seen as being more important, and therefore that has more impact, might also be mentioned more frequently during the group discussion. The latter possibility seems more likely. We found no evidence that any cue tended to have more of an impact on the group judgments that followed group discussions during which that cue was mentioned. On average, the within-group interaction terms that tested whether each cue had more of an influence on the group judgment when it was pooled were not significantly different from zero.

In addition, nonsignificant interactions between cue distribution and pooling show that the effect of cue distribution on cue impact did not depend on the level of pooling. Whether groups pooled a particular cue relatively frequently or relatively infrequently, the difference due to differential distribution was the same. Thus, the common knowledge effect was not, for example, produced by those groups that tended not to pool a cue.

In total, the content analysis suggests that, at least in the present research paradigm, differential pooling rates do not mediate differential cue impact. Most of the cues did tend to be pooled more frequently when they were shared and also when they had more impact on the group judgments. However, these effects both occurred when the other was controlled statistically. In addition, the relationship between pooling and impact appeared to occur only, on average, between groups and not judgment to judgment within groups.

Although we are not critical of Stasser's methods (indeed, his research was the inspiration for our own), we highlight one advantage of our application of social judgment theory statistical models (Hammond, Stewart, Brehmer, & Steinmann, 1986). In our study, we are confident that the primary mediator of information distribution effects (e.g., the common knowledge effect) is the differential impact of common and unshared information on individual judgments. Stasser's interpretation of behavior in his experimental task in terms of differential impact on the probability that information will be "pooled" in group discussion is most likely correct. Using his group choice task, it might be possible to distinguish between mediation through members' judgments and mediation through evidence pooling. However, we feel that our own paradigm presents a clearer picture of the relative importance of the two influences.

As formulated, the common knowledge effect hypothesis is concerned only with the general effect of information distribution on group judgments. In the present study, however, subjects were first asked to make individual judgments about each target student and then to come to a consensus judgment with the other group members. A simple combination of individual judgments into a group judgment, such as averaging, would result in the common knowledge effect. If group members simply combine their individual judgments into a group judgment, without any further utilization of the information, those cues that have an influence on more of the individual judgments will have more of an influence on the group judgment. On the other hand, the presented information might have an additional effect on group judgments over and above its effect on the individual group members' judgments. That is, the information could have an additional impact on the group judgments, controlling for the impact of the three members' judgments on the group's judgments. Moreover, the additional effect of a particular cue could depend on the number of group members who share that cue before the group discussion. For example, if the pooling of unshared information is affecting the group judgments, then the unshared cues should have more of an effect, when controlling for individual judgments. The groups should adjust their judgments in the direction of previously unshared information, but not in the direction of information that was fully shared. When unshared information is discussed, it provides information that is not already contained in all three member judgments, and therefore that information should have an additional impact on the group judgment. According to the common knowledge effect hypothesis, any such adjustment is not strong enough to overcome the initial bias. On the other hand, the common knowledge effect could be even stronger than originally formulated. The unequal distribution of information could have biased the group judgment even more than would be expected with a linear combination of the member judgments. For example, if a particular cue has an impact on the group discussion every time that it is pooled, and if the cue is pooled more often when it is fully shared than when it is unshared, then that cue would have the greatest additional impact on group judgments in the fully shared condition. Thus, differential pooling rates could augment the initial bias that is due to the impact of information on the individual members' judgments.

In this grade judgment task, none of the cues had more than a slight influence on the group judgments when the members' judgments were partialed out. The average regression weights were not reliably greater than zero for any of the cues, except for the attendance cue. In addition, there were no reliable linear effects of condition on the unmediated influence of the cues. In general, level of prediscussion information sharing does not affect the influence of information on group judgment, controlling for members' judgments. The regression analyses do not provide support for either of the possible effects that were discussed in the previous paragraph. Again, the evidence suggests that information pooling does not play a significant role in the group judgments.

In some ways, the grade judgment task kept the impact of information pooling modest. Because there was no accuracy feedback and no demonstrably correct solution (Stasser & Stewart, 1992), the group members had no way of knowing whether they could improve their accuracy by adjusting their group judgments to reflect pooled information. In addition, the structure of the task encouraged anchoring on the members' judgments. The group members made their individual judgments immediately before they began discussing each to-be-judged case. Moreover, the members made their judgments on the same "commensurate" scale as they used for their group judgment. All of the groups placed considerable weight on the member judgments during discussion. All three group members almost always pooled their member judgments during the discussion of each to-be-judged case. Moreover, as can be seen in the example transcripts, this opinion pooling usually occurred near the beginning of each discussion. Despite task instructions that emphasized the pooling of unshared information and group judgment accuracy, the groups in the present study tended to emphasize the information provided by the judgments of their constituent members.

Future research might investigate the generality of the common knowledge effect across different judgment tasks. For example, groups might be asked to make judgments for which there are demonstrably correct answers. They might also be provided with feedback concerning the accuracy of each of their judgments. In addition, the task could be modified to reduce the tendency for groups to anchor their judgments on the judgments of their members. Member judgments of all cases could be collected before group discussion of the first case. Alternatively, members and groups could render their judgments on two different, less commensurate response scales. Together or separately, these modifications might increase the influence of information pooling on the group judgments, relative to the impact of the member judgments. However, we emphasize again the similarity of the judgment task used in the present study to some real-world judgments. In a student petition hearing committee, for example, committee members would be likely to render their own implicit or explicit judgments as a case is presented, immediately before group discussion of that case. Moreover, both the members and the committee probably make their judgments on the same or similar response scales. Both members and groups might judge whether the committee should recommend that the petition be accepted or rejected. Although modifications of the judgment task might increase the impact of pooled information on the group judgments, such findings would not invalidate the results of the present study.

One possible explanation for the common knowledge effect is that subjects consciously determine that certain pieces of information are more important, because they are shared for every judgment case. However, even after making a large number of grade judgments, the individual group members are not able to assess accurately how they are using the provided information. In their subjective weightings of the six types of presented information, group members rank the influence of five of the six cues out of order, compared with the actual weights determined by the regression analysis for their group judgments. In addition, they give anxiety and workload in other classes subjective weightings that are a good deal higher than they should be. In reality, these cues, especially anxiety, had little influence on the groups' grade judgments. The subjects had a poor subjective insight into how they actually were using the information to come to a judgment about a case.

The between-conditions differences in judgment accuracy are puzzling on the surface. The groups in Condition 2 were considerably less accurate than groups in any other condition. Yet in these groups, the test percentile and attendance percentage cues were fully shared and thus had a large impact on the group judgments. However, for the 32 to-be-judged cases, attendance percentage was not a particularly good predictor, whereas high school grade point average and enjoyment, both unshared in Condition 2, were good predictors. Those two good predictors had a high impact among groups in Condition 1 (the impact of enjoyment being greatest when shared by two group members), and therefore those groups were the most accurate. Thus, the inability of groups to utilize unshared information did result in reliably different accuracies between the different experimental conditions, even though the results were initially perplexing.

The group judgments were not reliably more accurate than were the averages of the individual judgments in any of the conditions (Hastie, 1986; Hill, 1982). Allowing the group to come to a consensus judgment did not improve judgment accuracy, even though the individual judgments were underinformed because of the differential distribution of information. This result was expected for groups in the control condition, in which all group members made fully informed judgments before the group discussion. Groups in Condition 2, at the very least, could have been more accurate had they effectively pooled and weighted the cues that were more predictive of grades in the 32 to-be-judged cases.

What, then, can we conclude about group process with the present group judgment task? Taken together, our results suggest that the groups in the present study tended to use an implicit or explicit averaging combination rule to join their member judgments into a consensus group judgment. They did so even when they discussed judgment-relevant information that had influenced only 1 of the 3 averaged member judgments. That is, the groups did not adjust their judgments to reflect unshared information that was pooled during the group discussion. Our analyses suggest that information discussion in these groups functioned primarily as justification by members for their judgments, rather than as input into the judgment policies of the groups. The primary information that served as input into the groups' judgments was the judgments of their members.

The present study shows the inability of small groups to compensate for the differential distribution of judgment-relevant information among group members. The differential sharing of information affected the judgment policies of the groups. When individual group members pooled their information, the groups failed to take advantage of the cues that were unshared. The group judgments were not any more accurate than the average of the individual judgments, even though the groups were in possession of more information than were any of the individuals. Placing individuals in groups did not result in better judgments than would have been obtained by simply averaging together the judgments of the same individuals.

References

Brehmer, B., & Hammond, K. R. (1973). Cognitive sources of interpersonal conflict: Analysis of interactions between linear and nonlinear cognitive systems. *Organizational Behavior and Human Performance, 10,* 290–313.

Brehmer, B., & Joyce, C. R. B. (Eds.). (1988). *Advances in psychology: Vol. 54. Human judgment: The SJT view.* Amsterdam: Elsevier.

Brown, R. (1986). *Social psychology, the second edition.* New York: Free Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley: University of California Press.

Burnstein, E. (1982). Persuasion as argument processing. In H. Brandstatter, J. H. Davis, & G. Stocker-Kreichgauer (Eds.), *Group decision making* (pp. 103–124). San Diego, CA: Academic Press.

Deutsch, M., & Gerard, H. B. (1955). The study of normative and informational social influence on individual judgment. *Journal of Abnormal and Social Psychology, 51,* 629–636.

Hammond, K. R. (Ed.). (1966). *The psychology of Egon Brunswik.* New York: Holt, Rinehart, & Winston.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. (1986). Social judgment theory. In H. R. Arkes & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 56—76). New York: Cambridge University Press.

Harmon, J., & Rohrbaugh, J. (1990). Social judgment analysis and small group decision making: Cognitive feedback effects on individual and collective performance. *Organizational Behavior and Human Decision Processes, 46,* 34–54.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology, 46,* 44–56.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Decision research* (Vol. 2, pp. 129—157). Greenwich, CT: JAI Press.

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

Judd, C. M., & McClelland, G. H. (1989). Data analysis: A model-comparision approach. San Diego, CA: Harcourt, Brace, Jovanovich.

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83,* 602–627.

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 5–27). Chicago: University of Chicago Press.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.

Slovic, P., & MacPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance, 11,* 172–194.

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

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology, 63,* 426–434.

Stasser, G., Taylor, L. A., & Hanna, C. (1989). Information sampling in structured and unstructured discussions of three-and six-person groups. *Journal of Personality and Social Psychology, 57,* 67–78.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48,* 1467–1478.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53,* 81–93.

Stewart, T. R. (1988). Judgment analysis: Procedures. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The social judgment theory view* (pp. 41–74). Amsterdam: North-Holland.

Received August 28, 1992
Revision received May 4, 1993
Accepted May 11, 1993 ■