Information Exchange in Group Decision Making:

The Hidden Profile Problem Reconsidered[1]

John P. Lightle
Department of Economics, Florida State University, Tallahassee, FL 32311, jlightle@fsu.edu

John H. Kagel
Department of Economics, Ohio State University, Columbus, OH 43210, kagel.4@osu.edu

Hal R. Arkes
Department of Psychology, Ohio State University, Columbus, OH 42310, arkes.1@osu.edu

Group decision making provides a mechanism for channeling individual members' knowledge into productive organizational outcomes. However, in hidden profile experiments in which group members have common information favoring an inferior choice, with private information favoring a superior choice, groups typically choose an *inferior* alternative. We report a hidden profile experiment where we induce homogenous preferences over choice characteristics and provide financial incentives so that the common purpose assumptions of the model hold more completely than in past experiments. Nevertheless, groups continue to choose an inferior alternative most of the time. These failures primarily result from mistakes in recalling information, in conjunction with the fact that mistakes in recalling common information (which favors an inferior candidate) are typically corrected, while mistakes in recalling the private information needed to uncover the hidden profile cannot be corrected. As such the dismal performance of groups in pooling the information needed to identify the superior option primarily result from the structure of the problem rather than deficiencies in how groups share and process information. The discussions necessary to resolve mistakes in recalling common information also help to

[1] This research was supported by a grant from the Program in Decision, Risk, and Management Science at the National Science Foundation (SES-0452911). The authors would like to thank Susan Rose, Leonard Kiefer, David Cooper, and helpful seminar participants at Ohio State University, the International Meeting of the Economic Science Association in Atlanta, and the comments of the Associate Editor and two anonymous referees. The usual caveat applies.

explain the oft-noted fact that groups spend a disproportionate amount of time discussing common information.

Key words: group decision making; information exchange; hidden profile

1. Introduction: The idea that group decisions are more informed than individual decisions has considerable intuitive appeal. Groups bring together individuals with unique perspectives, information, and expertise which, if pooled efficiently, should be able to achieve superior outcomes compared to individual decisions, or the average of a set of individual decisions. Nevertheless there is a relatively long line of research that challenges this intuition, showing that in the presence of a "hidden profile," groups consistently fail to pool the available information, achieving distinctly sub-optimal decisions (Stasser and Titus, 2003). Given that many decisions within organizations are made by groups, these results are, to say the least, disappointing.

Consider a group of individuals faced with a small set of choices, with the task of reaching a decision on the alternative with the best set of characteristics. There are a finite set of characteristics common to all the choices, with at least one member of the group having information about every characteristic for each alternative. Information about these characteristics consists of either (i) common information available to *all* members of the group or (ii) private information that is known by a *single* member of the group. When a hidden profile is present the common information supports a suboptimal alternative. But if the entire group pooled its available information, and evaluated it in an unbiased way, they would prefer a different, superior alternative.

What the history of hidden profile experiments show is that groups typically fail to fully share their information, consistently failing to choose the optimal alternative. A number of reasons for this have been reported in the literature. One of the leading causes is that groups disproportionately discuss common information as opposed to the group members' private information (Stasser and Stewart, 1992). Other factors include information overload (too much information for individuals to remember) and biased recall that favors the alternative that each individual's pre-discussion information indicates is the best alternative (Stasser and Titus, 2003). Although the failure of team members to fully share information is not always detrimental to group performance, it often will be when a hidden profile is present.

The failure of groups to overcome the hidden profile problem is quite disturbing from the point of view of organizational behavior since it suggests that groups, although having access to all the relevant information to identify an optimal choice, fail to do so (Rulke and Galaskiewicz, 2000). Results from hidden profile experiments have been extensively cited in the management science/organization behavior literature. Among other things, the results have been used to call into question the superiority of diversity in organizations (Klein and Harrison, 2007), as well as suggesting the optimal level of diversity in organizations (Harrison and Klein, 2007). The results have been used to call into question the nature of group interactions, with a number of researchers pointing out the negative implications of failures to solve the hidden profile problem (e.g., Dennis, 1996; Thomas-Hunt, Ogden, and Neale, 2003) along with suggestions as to how to help overcome the problem (Weber and Donahue, 2001; Brodbeck et al., 2007).

Failures of group members to effectively pool information to discover a hidden profile occur in the context of a common purpose problem, where the incentives of members of the group are aligned both in terms of their preferences over the characteristics of each choice option, and the weights attached to these characteristics. However, the literature indicates a disturbingly large deviation from this ideal; for example, when pre-discussion information was distributed in such a way as to establish consensus in favor of an inferior candidate, as many as 33% of the subjects preferred another candidate (Stasser and Titus, 1985). In response, Experiment 1 replicates a classic hidden profile experiment. But rather than employ descriptions of characteristics that subjects typically view as good, bad or indifferent, we *assigned* ratings to each of the characteristics and provided explicit monetary incentives for choosing the optimal alternative. This is a standard strategy in experimental economics and indeed served to create substantially more homogenous preferences than reported in the original experiment. Nevertheless, this failed to make a dent in the groups' ability to uncover the hidden profile, as groups still identified the optimal candidate *less* than would be predicted by simply guessing. Analysis of this data showed that even if subjects were willing and able to share all the information that they remembered, without bias, they could not remember enough information to consistently uncover the hidden profile.

In response to this Experiment 2 employed the same procedures as Experiment 1 but reduced the set of characteristics to a manageable level so that given the average rate of recall, groups should have been able to uncover a little over 75% of the hidden profiles assuming subjects were willing and able to share all the information they could remember without bias. Here too groups fell well short of what was predicted as they uncovered only 35% of the hidden profiles. We explore three factors that might be responsible for this shortfall: heterogeneity in the amount of information recalled by subjects and therefore by groups, biases in what information is recalled, and errors in information recalled (e.g., a given alternative's characteristic is mistakenly recalled as positive when in fact it was negative or neutral). We find that the first two factors, by themselves, play a minor role in the failure to uncover hidden profiles. But *mistakes* in recalling information, in conjunction with the underlying *structure* of the problem, account for most of the failures compared to what was predicted. The reason for this is that in a hidden profile problem, mistakes in recalling common information can--and typically are--corrected, with this information favoring the suboptimal candidate. However, mistakes with respect to private information cannot, by definition, be corrected, and correct private information is necessary to uncover the hidden profile. We label the impact of these mistakes the "group correction factor," as correction of mistakes about common information naturally biases choices in favor of the suboptimal choice when a hidden profile is present.

Mistakes in recalling information have not been discussed before as a factor contributing to groups' failure to discover hidden profiles. In addition, the fact that most of the mistakes with respect to common information are corrected in the course of group discussions helps to explain the otherwise puzzling finding that groups spend a disproportionate amount of time discussing common as opposed to private information. However, this is exactly what one would expect in the presence of errors and the natural tendency of group discussions to focus on cleaning up those inconsistencies.

Experiments on the hidden profile problem are part of a broader experimental literature on information aggregation, or the lack thereof, in groups. Work on asset markets (reviewed in Sunder, 1995) shows that markets are capable of disseminating and aggregating private information, but this will not be achieved instantaneously or without replication, and will fail in some environments. Work on information cascades shows that a sizable minority of decision sequences result in "reverse cascades," where agents either start a chain of *incorrect* decisions due to initially misrepresentative signals, or fail to initiate correct cascades even when no misrepresentative signals are present (Anderson and Holt, 1997; also see Goeree et. al, 2007 for how "errors" in Bayesian updating with long sequences should eventually switch to the correct state). Blinder and Morgan (2005) show that groups make somewhat better decisions on average than individuals, but that the demand for additional information does not depend on whether the decision is made by an individual or a group. In decisions regarding *risky* choices, groups tend to decide differently than individual members would on their own, sometimes resulting in making *riskier* choice, and at other times resulting in greater *restraint* in risk taking (Davis et. al, 1992). Thus, the collapse of information aggregation documented in the hidden profile problem is not unique.[2] What our experiment does is to identify a previously undiscovered factor which may be responsible for the dismal performance of groups in discovering hidden profiles, a factor that has nothing to do with deficiencies in how groups process information or make decisions, but is rather a structural element underlying the presence of a hidden profile.

2. Experiment 1*:* Experiment 1 is designed to evaluate the effect of explicit incentives on the ability to overcome the hidden profile (HP) problem. We replicate the original HP experiment (Stasser and Titus, 1985), including some of the profiles they employed, where each of three candidates for political office had 16 relevant characteristics, with each subject provided with information about 10 of the characteristics. Two treatment conditions were employed: One in which groups were permitted to take

[2] Hidden profile research is also related to the literature on information aggregation, in which biases have been documented due to the presence of public information that introduces correlations between the input of multiple members of a market or other group (Chen, Fine, & Huberman, 2004).

their information about candidate characteristics into the caucuses (the "perfect recall" condition), the other in which they were not, as in the typical HP experiment (the "imperfect recall" condition). The first treatment is a calibration exercise to determine if explicit incentives, in conjunction with perfect recall, will achieve near 100% group agreement in favor of the optimal candidate. Given that the perfect recall treatment achieves this, the second treatment is designed to determine the impact of our experimental procedures on the ability to discover HPs.

2.1 Method: Two experimental sessions were conducted, each of which employed 12 undergraduate students attending The Ohio State University, who responded to e-mails to participate in a voting experiment in which they would earn money as a result of their choices. For each session recruitment e-mails were sent to approximately 100 students enrolled in undergraduate economics classes.

Each experimental session had five "elections" with subjects randomly re-paired into caucus groups of four subjects each. The first three elections employed the imperfect recall treatment, followed by two rounds of perfect recall. This order was dictated by the fact that had we used the reverse order, it may well have alerted subjects to the fact that their own information favored an inferior candidate, thereby biasing the imperfect recall treatment in favor of discovering the HP.

At the beginning of the session instructions were read aloud with subjects having copies to follow along with. Prior to each round subjects were given an information sheet which listed the ratings of the three candidates on 10 out of the 16 characteristics. The ratings were designed to establish a HP, with round 2 replicating Stasser and Titus's (1985) Unshared/Conflict HP treatment, and rounds 3 and 5 replicating their Unshared/Consensus HP treatment.

	Candidate A	Candidate B	Candidate C
Round 1- Random			
Positive	6[random]	5[random]	5[random]
Neutral	8[random]	5[random]	6[random]
Negative	2[random]	6[random]	5[random]
Round 2 – Conflict			
Positive	8 [2]	4 [common]	4 [common]

Neutral	4 [common]	8 [6,4]	8 [4,6]
Negative	4 [common]	4 [0,2]	4 [2,0]
Round 3 – Consensus			
Positive	8 [2]	4 [common]	4 [1]
Neutral	4 [common]	8 [5]	8 [common]
Negative	4 [common]	4 [1]	4 [1]
Round 4 – Random			
Positive	9[random]	6[random]	5[random]
Neutral	1[random]	5[random]	7[random]
Negative	6[random]	5[random]	4[random]
Round 5 – Consensus			
Positive	8 [2]	4 [common]	4 [1]
Neutral	4 [common]	8 [5]	8 [common]
Negative	4 [common]	4 [1]	4 [1]

Table 1: Total Characteristics of Candidates and Pre-Discussion Distribution of Information in Experiment 1: Rounds 1-3, Imperfect recall; Rounds 4-5 Perfect Recall

Table 1 summarizes how the pre-discussion information was distributed. When “common” is inside the brackets, it indicates that each member of the caucus group got pre-discussion information about each of the characteristics; e.g., in Round 3 candidate A had 4 negative characteristics with each caucus member receiving information about *all* 4. When numbers appear in the brackets it indicates the number of pieces of private information each caucus member was provided; e.g., in Round 3, Candidate A has 8 positive characteristics, with each caucus member having pre-discussion information about 2 of the 8 characteristics. When a comma appears inside a bracket it means that the two of the caucus members got information about the first number of characteristics, with the other two getting information about the second number; e.g., in Round 2, there were 4 negative characteristics for Candidate B with two members of each caucus group getting no negative information about B and the other two each getting 2 of B’s 4 negative characteristics. When “random” appears inside the bracket, information was distributed randomly so as to ensure a HP (details are provided in the appendix). The random treatment was employed to ensure that subjects would not recognize any systematic patterns to their private information across rounds.

In Table 1 Candidate A is always the optimal candidate, with Candidate B or C the sub-optimal candidate preferred on the basis of pre-discussion information. Candidate letters changed between rounds (e.g., D, E, and F were used in Round 2) with the letter for the optimal candidate randomized among the three choices. The names of the characteristics remained the same across rounds. Round 2 is designed to get two of the four caucus members to prefer B on the basis of their pre-discussion information, with the other two preferring C. Rounds 3 and 5 are intended to induce a consensus in favor of B based on pre-discussion information. Rounds 1 and 4 were designed so that at least 3 of the 4 caucus members would strictly prefer a sub-optimal candidate, with the remaining member approximately indifferent between the optimal candidate and one of the sub-optimal candidates. The relaxed criteria for a HP in these rounds should have made finding the optimal candidate no harder than in the other three rounds.

In the Imperfect Recall treatment, subjects were given three minutes to study their information sheets about the candidates' characteristics, after which the sheets were collected. Immediately following this, subjects began to fill out their questionnaires for the round, which asked them to write down every rating for each candidate that they could recall within a two-minute period, and to indicate their preferred candidate. Subjects were paid $0.25 for every correct piece of information, relative to the information they received, for one randomly selected round. However, to prevent rampant guessing, if a subject's questionnaire had more than 3 incorrect answers, they earned $0.00, regardless of the number of correct pieces of information recalled. In the Perfect Recall treatment, subjects brought their pre-discussion information sheets into the caucuses, but indicated their preferred candidate prior to caucusing.

After the questionnaires were filled out subjects were (randomly) assigned to caucus groups, with each group meeting in a designated caucus area to discuss the candidates. Subjects had a checklist of all 16 characteristics and a reiteration of the formula for computing the monetary payoff for each candidate at the bottom of the checklist. Groups had 5 minutes to discuss the candidates after which they were asked to vote by secret ballot. A candidate needed 3 out of 4 votes, otherwise no candidate was declared a winner, yielding a payoff of $0 for that round. (All groups managed to select a candidate, typically unanimously.) Following the vote subjects were told the 16 characteristics of the chosen candidate (positive, negative or neutral) along with the monetary payout for that candidate. Subjects received no information about the characteristics of the other candidates to prevent them from learning that their pre-discussion preferred candidate was sub-optimal.

Payoffs for characteristics were computed as follows:

Payoff = (# of positive ratings)*$1.00 + (# of neutral ratings)*$0.00 - (# of negative ratings)*$0.50.

This provides an obvious financial incentive for choosing the optimal candidate. Each characteristic had a natural meaning presumably relevant for choosing between elected officials – for example, age, work experience, military service, etc., with nothing other than the rating for the relevant characteristic reported (positive, neutral or negative). Subjects were told that each of their ratings of a candidate's characteristics was identical to that of other subjects and that individually they did not necessarily have all the relevant information about all of the candidates' characteristics, but that the value of any candidate was a function of *all* of the candidate's characteristics. The checklist allowed groups to record the content of their discussions regarding payoff relevant information and to compute the payoffs for each candidate.

At the end of the session subjects received a cash payment equal to the sum of their earnings from the caucuses plus their earnings from the randomly selected pre-caucus questionnaire along with a $2 participation fee. Total earnings averaged $20.83 for a session that lasted approximately 1.5 hours.[3]

2.2 Results: The top part of Table 2 shows pre-discussion preferences, which are quite strong in favor of one or the other of the sub-optimal candidates, just as the pre-discussion information structure was intended to produce. Evaluating the effects of explicit incentives, 90% (43 out of 48) of the subjects preferred the sub-optimal candidate B in the Consensus rounds (3 and 5). This compares to 67% in favor of the sub-optimal consensus candidate in Stasser and Titus (1985). Further, the conflict treatment,

---

[3] A copy of the full set of instructions including the questionnaire and checklist of candidate characteristics is provided in the appendix.

intended to create a 50-50 split between the two sub-optimal candidates does just that. These results support the goal of establishing clear and unambiguous preferences through explicit incentives and inducing the desired preferences over candidate characteristics, establishing a much closer correspondence to the common purpose norm than in previous studies.[4]

Pre-Discussion Preferences†			
Round	Candidate A	Candidate B	Candidate C
1- Random	.29	.46	.25
2 – Conflict	.00	.50	.50
3 – Consensus	.00	.92	.08
4 – Random	.13	.38	.50
5 - Consensus	.08	.88	.04

Post-Discussion Preferences††			
Round	Candidate A	Candidate B	Candidate C
1- Random	.33	.67	.00
2 – Conflict	.00	.83	.17
3 - Consensus	.00	.83	.17
4 - Random*	1.00	.00	.00
5 - Consensus*	1.00	.00	.00

*Perfect recall treatment.
† Twenty-four individuals choosing in each round.
†† Six groups choosing in each round.

Table 2: Pre- and Post-Discussion Preferences over Candidates in Experiment 1

[4] There is clearly some loss of control/confusion for some subjects given the failure to obtain 100% preference for the consensus candidate in Round 5 (3 out of 24 subjects). Round 4 was designed to make half the caucus members prefer B and the other half prefer C. However 3 out of 24 subjects chose A. One possible reason behind these discrepancies is that subjects were not paid contingent on the accuracy of the pre-discussion preferences while being paid for all other elements of the task.

(fraction preferring each candidate)

The bottom half of Table 2 shows the caucus group selections. With perfect recall (rounds 4 and 5) the HP was always identified (12 of 12) which is significantly better than a random guess norm.[5] These results also support the notion that our procedures established a close correspondence to the common purpose norm assumed in the HP problem. In contrast, in the Imperfect Recall treatment, 11.1% of the groups (2 of 18) identified the optimal candidate, which is significantly less ($p < .05$) than with random guessing.

2.3 Discussion: Introducing explicit incentives into a HP problem, while serving to better control preferences, did not do much to alleviate the HP problem in the imperfect recall treatment. In what follows we develop an idealized model of group discussions to serve as a reference point against which to identify the relative importance of different factors, such as biased recall, memory limitations, etc., contributing to the failure to discover HPs. Ours is not the first choice model applied to the HP problem (see, for example, Stasser and Titus, 1987; Larson, et al., 1994; Stasser, 1988; Stasser and Taylor, 1991) and has a number of ingredients in common with these earlier modeling efforts. What is different is our goal of developing a simple, *idealized* model of group discussions which, if adhered to, would assure identifying HPs, as opposed to a descriptive model of the discussion process. Using this as our reference point we look for breakdowns in behavior, relative to the model, to identify their potential impact on the failure to discover HPs. As such in our model there are *no* biases in recall and subjects mention and fully account for *all* the information they can remember, factors whose absence the literature indicates are important elements behind the failure to discover HPs.

The idealized model assumes the following: (1) Every characteristic is equally likely to be recalled, with the same probability p, so there is no biased recall in favor of pre-discussion preferences; (2) Every subject has the same p, so there are no group members who are ex-ante better or worse at recalling information (no weak links); (3) Recalling a single piece of information is an independent event

[5] Ours is not the only experiment to employ a perfect recall treatment (e.g., Hollingshead, 1996). However, ours is the first we know of in which even with perfect recall the hidden profile was detected 100% of the time.

(Bernoulli trial); (4) Subjects never make an error in recall; they either recall information accurately or do not recall the information; (5) All information that is recalled is shared with the other group members: (6) Groups select the candidate that maximizes their expected payoff conditional on the entire set of information one or more caucus member recalls. Under these assumptions the probability that a piece of private information is recalled in caucus is p, while the probability that a piece of common information is recalled is $1 - (1-p)^4$, as the latter only fails to be recalled when all four group members independently fail to recall it.

Given the payoffs for characteristics and the distribution of information in round 3 of the incomplete information treatment, p must be large enough for the following inequality to hold in order for the expected payoff of the optimal candidate to be larger than the expected payoff of the sub-optimal candidate[6]:

$$4(-\$0.50)(1-(1-p)^4) + 8(\$1.00)p > 4(\$1.00)(1-(1-p)^4) + 4(-\$0.50)p$$

This inequality holds if and only if p is larger than 0.58, so that group members would each need to recall slightly more than half of their pre-discussion information (18 out of the 30 candidate characteristics) without bias, or mistakes, to have a better than 50% chance of identifying the HP; i.e., the 58% recall rate is a *necessary* condition for the group to have a better than 50% chance of discovering the HP.[7] It is not a *sufficient* condition since all the other assumptions of the benchmark model would also need to be satisfied to have a 50% chance of discovering the HP.

---

[6] The left hand side of the inequality consists of the 4 pieces of common information regarding the negative characteristics for the optimal candidate (A) plus the 8 positive characteristics which are all private information. The right hand side of the inequality consists of the 4 positive characteristics (all common information) for the sub-optimal candidate B along with the 4 negative characteristics which are private information.

[7] The model allows us to create a function, f(p), which gives the predicted probability of HP detection given p. To derive this let x be a vector of random variables, where x_i equals 1 if the i^{th} piece of information was recalled by at least one caucus member, 0 if it was not. Let $g(x) = 1$ if the realization x results in the group discovering the HP, $g(x) = 0$ otherwise. Now, given p, let $H_p(x)$ be the probability that x is realized. Then f(p) is the probability that a group will discover the HP, which is given by the sum of $H_p(x)$ over all x such that $g(x) = 1$. See Experiment 2 for further discussion.

The estimated recall rates based on the pre-discussion information sheets averaged 33.2%, much lower than the 58% needed under the assumptions of the model to have a 50% chance of discovering the HP. Thus, the failure to discover HPs in Experiment 1 can be attributed to information overload. However, this does little to explain the persistence of the HP problem as it is present in experiments in which there are substantially fewer characteristics associated with the choices in question (e.g. Stasser and Stewart, 1992; Stasser et al., 1995; Fraidin, 2004). We therefore ran Experiment 2 in order to study the frequency and causes of the HP problem when information overload is less likely.

3. Experiment 2:

3.1 Method: *S*ubjects were 32 undergraduate students enrolled in economics classes at The Ohio State University who responded to e-mails asking them to participate in a voting experiment in which they could earn money. Three sessions were conducted with 12 subjects in sessions 1 and 2 and 8 subjects in session 3.

Procedures were essentially the same as in Experiment 1 except that the total number of characteristics for each candidate was reduced to 7, with subjects receiving information about 4 of the 7 characteristics. With fewer characteristics, the time subjects had to look at their pre-discussion information was reduced to 2 minutes. Payoffs were changed to $2.00 for every positive rating, and -$1.00 for negative ratings, as this was necessary to keep incentives comparable to Experiment 1. Payoffs for correct questionnaire responses were increased to $0.50 each, with three or more mistakes penalized with zero payment. We dropped the perfect recall treatment having already demonstrated that with explicit payoffs and full information failures to discover HPs would be largely eliminated.

	Candidate A	Candidate B	Candidate C
Round 1 - Random			
Positive	3[random]	3[random]	3[random]
Neutral	3[random]	2[random]	1[random]
Negative	1[random]	2[random]	3[random]
Round 2 - Conflict			
Positive	4 [1]	2 [common]	2 [common]

Neutral	0	3 [1,2]	3 [2,1]
Negative	3 [common]	2 [1,0]	2 [0,1]
Round 3 - Consensus			
Positive	4 [1]	3 [common]	1 [common]
Neutral	0	0	6 [3]
Negative	3 [common]	4 [1]	0
Round 4 - Random			
Positive	4[random]	2[random]	2[random]
Neutral	1[random]	4[random]	4[random]
Negative	2[random]	1[random]	1[random]
Round 5 - Consensus			
Positive	4 [1]	3 [common]	1 [common]
Neutral	0	0	6 [3]
Negative	3 [common]	4 [1]	0

Table 3: Total Characteristics of Candidates and Pre-Discussion Distribution of Information in Experiment 2

The distribution of common and private information was designed to mimic the distribution in Experiment 1. We refer the reader to Table 1 for how to read the distribution of information in Table 3. The exact distribution of information in the random rounds is provided in the appendix.

3.2 Results:

Pre-Discussion Preferences[†]			
Rounds	Candidate A	Candidate B	Candidate C
1 - Random	.00	.45	.55
2 – Conflict	.00	.53	.47
3 - Consensus	.00	1.00	.00
4 – Random	.05	.75	.20
5 – Consensus	.11	.79	.11

Post-Discussion Preferences††			
Rounds	Candidate A	Candidate B	Candidate C
1 - Random	.25	.50	.25
2 – Conflict	.50	.25	.25
3 - Consensus	.25	.63	.13
4 – Random	.50	.38	.13
5 – Consensus	.25	.63	.13
Average	.35	.48	.18

† Thirty two individuals in each round.
†† Eight groups in each round.

Table 4: Pre-Discussion and Post-Discussion Preferences Over Candidates in Experiment 2 (fraction preferring each candidate)

Table 4 reports pre- and post-discussion preferences (top and bottom parts, respectively). In the Consensus treatment (rounds 3 and 5) 87% of pre-discussion preferences favored Candidate B, slightly less than the 90% reported in Experiment 1, but still well above the 67% reported in Stasser and Titus (1985). The Conflict treatment (round 2) comes very close to achieving the 50-50 split in pre-discussion preferences between sub-optimal candidates B and C intended. Even with the reduction in the number of characteristics the average overall rate of HP detection was only 35.0%, which is *not* significantly different from the random guess norm of 33.3%.[8] Further, as the table shows, there were no systematic changes in the frequency of HP detection across the 5 rounds.

There is strong evidence *against* random guessing as subjects made extensive use of the checklists in the caucuses: 60% of the checklists were filled out in their entirety, with no checklist missing more than 29% of the characteristics. Further, 93% of groups chose the candidate that had the highest (or tied for highest) payout based on the checklist information. Thus, with few exceptions, groups were using

[8] The average rate of HP detection across sessions was 26.7% in session 1 and 40.0% in sessions 2 and 3.

the checklists to aggregate information and choose the candidate that maximized payoffs given their collective recall.

The caucus groups chose the optimal candidate substantially more than individual subjects would have done based on their pre-discussion preferences, going from 5% to 35%.[9] This stands in marked contrast to most HP experiments, and quite unlike the results from Experiment 1. Groups are doing substantially better than simply averaging over their individual pre-discussion preferences. However, subjects could have earned close to 50% more had they been able to discover all the HPs - $17.50 versus $26.00 (excluding questionnaire responses and participation fee).

3.3 Factors Underlying Failure to Discover Hidden Profiles*:* In what follows we employ the reference point model to identify factors responsible for the failure to detect more HPs. First, on average subjects *correctly* remembered 84.9% of the pre-discussion information they had been given, a substantially higher percentage than in Experiment 1. Subjects *incorrectly* recalled 12.7% of the pre-discussion information provided, and did not recall 2.4% of the information. If we ignore the information that is recalled incorrectly, and assume instead that it is simply not recalled, the 84.9% correct information recall rate together with the idealized model generates a predicted HP discovery rate of 78.0%. This is well *above* the 35.0% actually detected ($p < .01$). Figure 1 illustrates the model's prediction of HP discovery as a function of subjects' average rate of remembering, along with a simulated 95% confidence interval and the actual discovery rate under the assumptions of the reference point model.

[9] These percentages are significantly different at better than the 1% level using a binomial test, but suffer from a repeated measures problem for the pre-discussion preferences.

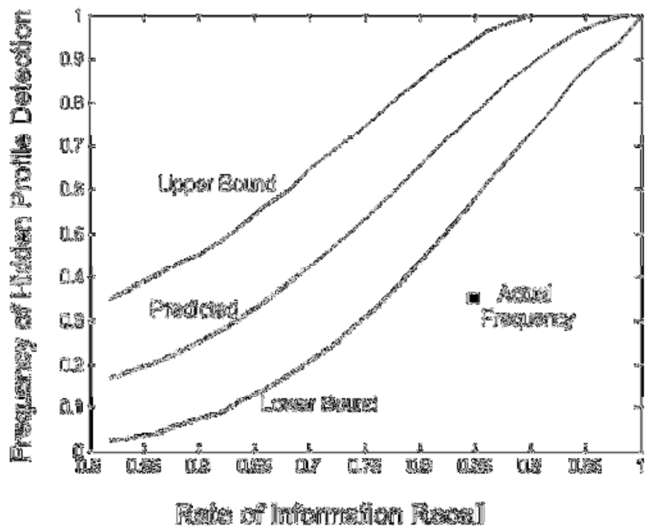


Figure 1: Predicted Frequency of Hidden Profile Detection in Experiment 2 with 95% Confidence Interval vs. Actual Frequency.

In what follows, we focus on three major failures, within the context of the model, potentially responsible for the shortfall in HP discovery: (1) heterogeneity within and between group in the amount of information recalled, (2) biased recall of information, and (3) the group correction factor. Although (2) is commonly reported in other HP experiments, (1) and (3) have not, to our knowledge, been discussed before. We conclude that the group correction factor, which results from *mistakes* subjects make in remembering characteristics, in conjunction with the structure of the problem, is the factor which is most responsible for the shortfall in HP discovery.

3.4 The Relationship of Information Recall and HP Discovery:

The reference point model assumes that any information which is recalled during the questionnaire phase will also be recalled and utilized in the caucuses where decisions are actually made. Therefore, the model makes a strong prediction that recalling more information leads to better decisions. However, using a probit regression we find that average number of correct pieces of information recalled has no significant effect on the probability of HP discovery.[10] Figure 2 reports the raw data underlying this result.

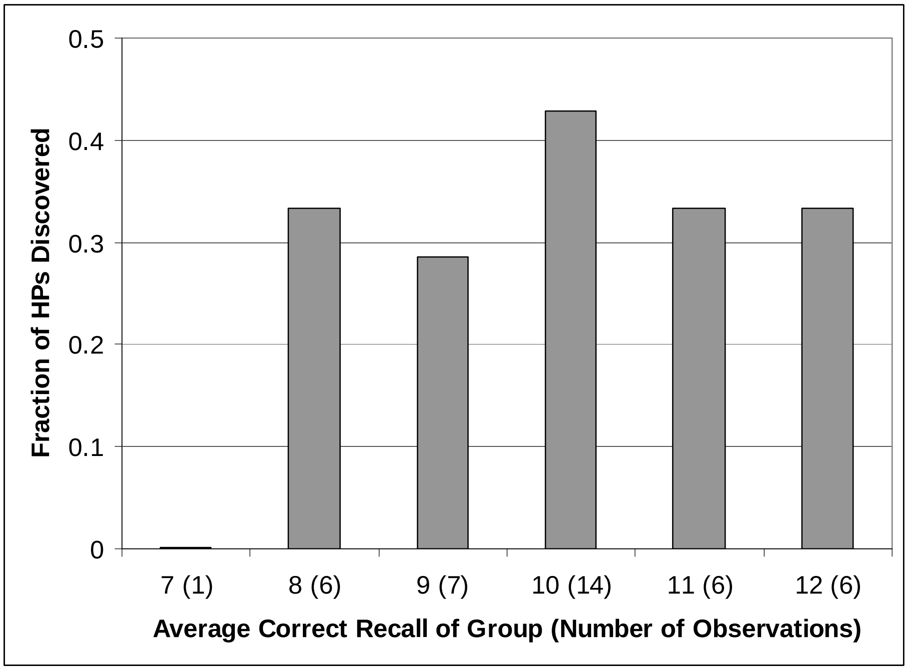


Figure 2: Group Performance as a Function of Average Recall

While these results seem to indicate that the reference point model is inherently flawed, there are three things to note before jumping to this conclusion: 1) The probit employs recall rates from the questionnaire phase of the experiment, but a piece of information only gets used if it is both recalled and mentioned in the caucuses. 2) We have ignored within-group heterogeneity in recall rates which is large and can adversely affect the number of HPs discovered. 3.) The probits only consider correct pieces of information recalled ignoring the potential effect of *mistakes* in what is recalled. The fact that 60% of the checklists were entirely filled out is evidence that subjects were willing to mention information. But the possibility remains that the information mentioned was inaccurate and these mistakes, combined with within-group heterogeneity, negatively affected the number of HPs discovered (and would not be captured in either the probit or Figure 2).

[10] The regression estimate is probit(p_{ir}) = -1.71 + 0.13 $AveRecall_{ir}$ where p_{ir} is the probability that the HP is detected by group i in round r, and probit(p) is the standard probit function, $AveRecall_{ir}$ represents the average number of correct pieces of information recalled by members of group

3.5 Heterogeneous Information Recall:

The reference point model assumes that every subject has the same ability to recall information and each piece of information is recalled with probability p, where p is a parameter of the model. Therefore, the model predicts that the number of pieces of information that a subject recalls (out of 12 possible) in any one round is a random draw from the binomial distribution with n equal to 12 and p equal to the probability parameter of the model.

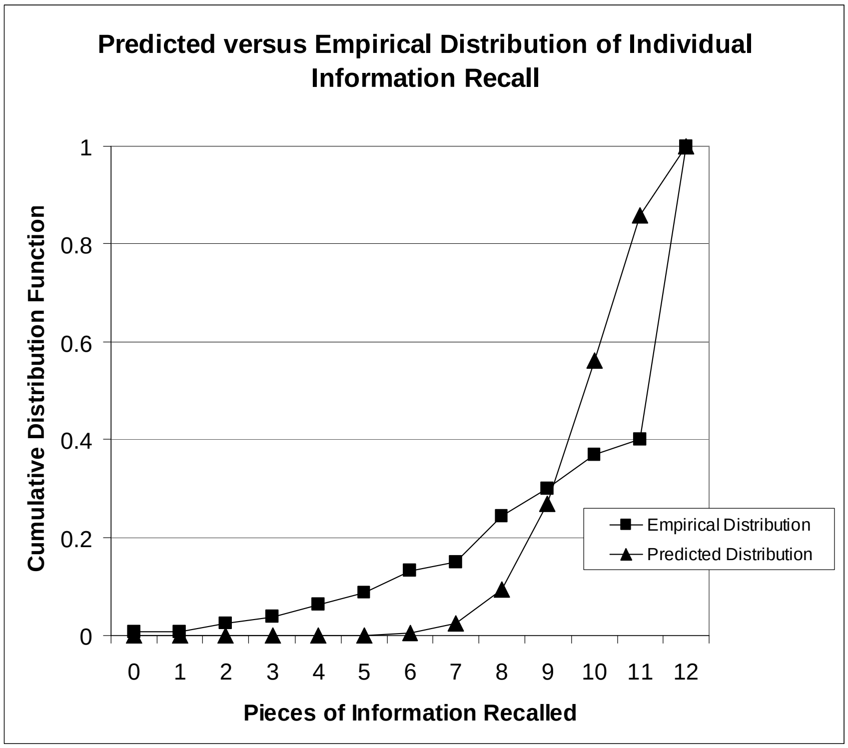


Figure 3: Predicted versus Empirical Distribution of Individual Information Recall

Figure 3 plots the distribution of predicted recall rates for individuals in the reference point model while setting p equal to the observed average recall rate of 84.9%. In addition, Figure 3 plots the empirical distribution of recall rates, which differs notably from the binomial distribution. In particular, there is a heavier left tail, meaning that a larger than predicted number of subjects had particularly poor recall. It is important to note that the HP task is one where each individual's contribution could be critical to choosing the optimal candidate, because unique information is precisely the information which favors

i in round r. The p-value on the coefficient of AveRecall is 0.45.

the optimal candidate. Thus the presence of the heavy left tail seen in Figure 3 could be responsible for the reduced number of HP discoveries. In other words, the actual within-group heterogeneity is larger than what is accounted for in the model and could be partially responsible for the shortfall in HP discovery.[11]

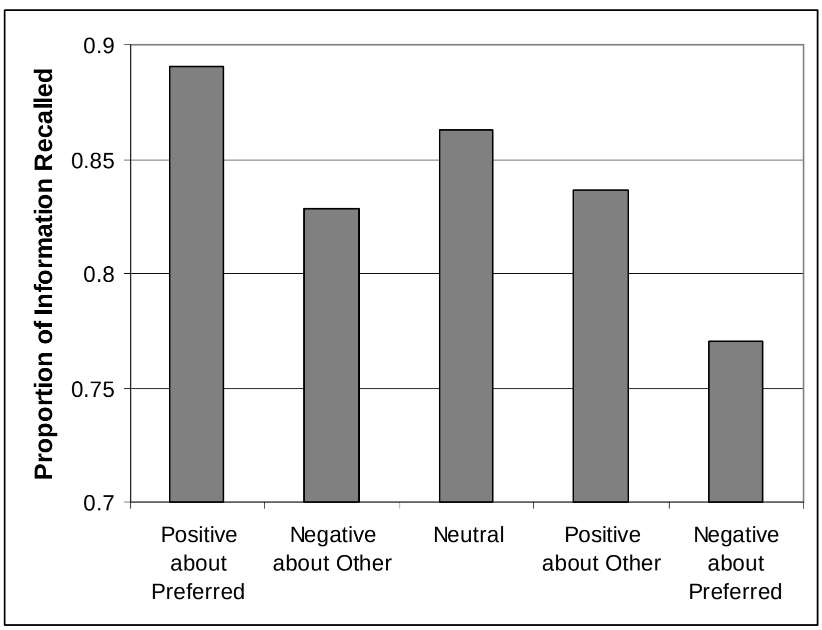
To determine the impact of heterogeneity on performance, we conduct simulations in which the precise set of information recalled by any subject in any round is randomly determined on the basis of the empirical distribution of Experiment 2 itself (given in Figure 3). This distribution gives rise to within-group and between-group heterogeneity which is, on average, equal to that of the experiment. The simulation determines which candidate each of the 40 groups would vote for based on what caucus members recalled, given the assumptions of the simple reference point model. In 10,000 simulated sessions, the average HP discovery rate was 77.0%, with a 95% confidence interval of between 65.0%-87.5%. Even the lower bound recall rate from the simulations, 65.0% is quite far away from the observed rate of 35%. As such we conclude that individual heterogeneity in information recall plays at most a modest role in the failure to identify HPs.

3.6 Biased Recall of Information*:* The reference point model's assumption that all pieces of pre-discussion information are equally likely to be recalled is not satisfied in our data. Figure 4 shows recall rates as a function of the type of information provided (all categories are in relation to the preferences that pre-discussion information was designed to induce): Positive information about the preferred candidate (PP), negative information about another candidate (NO), neutral information about any of the candidates (N), positive information about another candidate (PO), and negative information about the preferred candidate (NP). We fit the following random effects logit model to the data:

$$\text{logit}(p_{ijr}) = \underset{(0.18)^{**}}{2.08} + \underset{(0.22)}{0.30}\, PP_{ijr} - \underset{(0.19)}{0.30}\, NO_{ijr} \underset{(0.20)}{-0.27}\, PO_{ijr} \underset{(0.29)^{*}}{-0.70}\, NP_{ijr} + \alpha_i \qquad \text{Wald } X^2(4) = 14.5^{**}$$

[11] Similarly, the empirical between-group heterogeneity is slightly larger than what is accounted for in the model, but based on the results reported in Figure 2, we can ignore this effect.

Where $p_{ijr} = 1$ is the probability that subject i recalls information correctly with respect to category j in round r, and logit(p) = ln(p/1-p). The first four right hand side variables are dummy variables with value 1 for the information category in question (PP, NO, etc.) and zero otherwise. The variable α_i is a subject-specific effect to account for the heterogeneous ability of subjects to recall information. The baseline against which recall rates are evaluated is neutral information about candidates, which represents the intercept value. Standard errors are reported in brackets below the coefficient estimates, with two stars for a parameter whose value is significantly different from zero at $p < 0.01$, and one star at $p < 0.05$. The Wald statistic tests the null hypothesis that the different categories are remembered at the same rate, which can be rejected ($p < .01$). Negative information about one's pre-discussion preferred candidate (NP) is *least* likely to be remembered and is significantly less likely to be remembered relative to neutral information ($p < 0.05$). Re-specifying the model using positive information about the pre-discussion preferred candidate as the baseline, the null that NP = PP is rejected at $p < 0.01$.[12]



[12] It is worthwhile noting that mistakes in information recalled is biased in favor of a subject's pre-discussion preferred candidate: 16.5% of the possible negative or neutral information about a subject's pre-discussion preferred candidates is mistakenly recalled in *favor* of the preferred candidate compared to 7.3% for other candidates; 5.4% of the possible positive or neutral information about the preferred candidate is mistakenly recalled so as to *hurt* the preferred candidate as opposed to 9.0% for other candidates.

Figure 4: Rate of Information Recall as a Function of Consistency with Pre-Discussion Preference in Experiment 2

Biased recall in favor of the pre-discussion preferred candidate must reduce the probability of discovering the HP. To determine an upper bound for the marginal effect of the worst of the biases, the low recall rate for NP, we consider the following counterfactual: "What if negative information about the preferred candidate were recalled at the same rate as the average for all other information, with all other assumptions of the reference point model satisfied?" To do this we divide the relevant information of each group's decision into NP information, and all other information. For all other information, we assume that a piece of information was recalled if and only if it was actually recalled during the experiment. For NP information, we treat the recall of each piece of information as an independent random event such that with probability 0.849 the correct information is recalled, and with probability 0.151 nothing is recalled. Because the NP recall is random, the best we can do is calculate the expected number of times that a group which did not discover the HP would do so under the counterfactual[13].

The result is in an expected increase of 2.2 HPs being detected, which would increase the HP discovery rate from 35.0% to 40.5%. An increase in the HP discovery rate of 5.5% may seem substantial given the fact that we restricted our measurement of the effect of biased recall to NP information[14]. However, in addition to raising the correct recall rate of NP information from 75.7% to 84.9%, the counterfactual *eliminated* all the mistakes in recalling NP information, treating that information as not having been recalled. In Section 3.7, we will see that mistakes in information recall have a significant impact on whether or not a group discovers the HP. Since both raising the NP recall rate and eliminating these mistakes are responsible for the counterfactual's higher prediction of 40.5%, the effect of the lower

[13] Here we assume that any discrepancies that would emerge are resolved according to majority rule, and cases with no majority treated as no information. Ties among candidates are broken with equal probability.

[14] Had we included more than NP information in the counterfactual, the predicted increase in HP discovery would come closer and closer to the reference point model's prediction. In fact, if the counterfactual adjusted all information the way it does with NP, it would be identical to the reference point model, since all information would be recalled with probability 0.849, and there would be no mistakes. Therefore, it is necessary to look at the effect of biased recall in isolation to minimize interaction effects with other factors, in particular, the group correction factor.

recall rate for NP information alone must be smaller than this. In other words, 2.2 increase in HPs discovered is an upper bound on the marginal effect of the lower NP recall rate.

The biased recall of information here confirms one of the conclusions of the very first HP experiments: Subjects bias their recall of information in favor of their exogenously determined pre-discussion preferred candidate (Stasser and Titus, 1985). Our findings are consistent with prior research showing that people exhibit better recall for facts supporting their decisions (e.g., Dellarosa and Bourne, 1984). Biased recall is also consistent with prior research illustrating the benefits of within-group dissent; homogeneous groups whose members' opinions initially favor one candidate are more likely to prefer information supporting that candidate (Schulz-Hardt et al., 2000), with contrary information more likely to be disregarded.

3.7 Group Correction Factor: The assumption of the reference point model that information is always recalled correctly is not satisfied, with an average of 12.7% of the pre-discussion information recalled incorrectly; for example, a subject wrote down "positive" under candidate A for public service when the information was actually "negative." If a subject makes an error with respect to common information the group has an opportunity to correct that mistake. However, *there is no opportunity to correct mistakes with respect to private information*, because it is unique to an individual subject. Further, given the structure of HPs, common information favors a sub-optimal candidate, while private information favors the optimal candidate. As such, to the extent that mistakes with respect to common information are corrected by other members of the group, this favors a sub-optimal candidate being chosen. Mistakes with respect to private information, which by definition cannot be corrected, hurt the optimal candidate's chances of being selected, since no one else in the group can correct the mistake. Thus, a natural corrective force of group discussion, clearing up mistakes, biases the caucus groups to choose a sub-optimal candidate when a HP is present. We label this effect "the group correction factor."

To estimate the impact of the group correction factor we consider the following counterfactual: "What would happen if private information were corrected at the same rate as common information, given the average recall rates reported?" We examined all questionnaires to locate common information that was written down incorrectly. We cross referenced these cases with the group checklists to discover how many times the caucuses corrected the mistake, making it onto the checklist correctly. This happened 87.9% of the time (124 out of a total of 141 mistakes).[15]

Using the checklists filled out during group discussion, we can predict what a group would have done if an incorrect piece of private information was corrected at the same rate. Corrections of private information would never have resulted in a group selecting a sub-optimal candidate in place of the optimal candidate. In contrast, in 21 out of the 26 cases where the HP was not detected, there was incorrect private information which if corrected might have detected the HP. Applying the common information correction rate to these 21, would have resulted in an expected *increase* of 12.7 HP detections. Adding this to the number of HPs actually discovered (14), the groups would be expected to pick the optimal candidate 66.8% of the time.[16]

Further evidence of the importance of the group correction factor, in particular the effect of mistakes in private information, can be found between caucus groups. Using the group checklists of candidate characteristics, we estimated the following probit regression:

$$\text{probit}(p_{ir}) = \underset{(1.74)}{0.59} - \underset{(0.23)}{0.09}\,\text{ComCorr}_{ir} + \underset{(0.10)}{0.03}\,\text{PrivCorr}_{ir} \underset{(0.14)^{**}}{-0.28}\text{PrivWro}_{ir}$$

where p_{ir} is the probability that the HP is detected by group i in round r, and probit(p) is the standard probit function, ComCorr_{ir} and PrivCorr_{ir} represent the number of pieces of correct common and private information on i's check list, and PrivWor_{ir} is the number of pieces of incorrect private information on i's

[15] Correction rates were as follows: (i) 91.5% (108/118) when a majority had the correct answer, (ii) 73.7% (14/19) with the same number of correct and incorrect answers, and (iii) 50.0% (2/4) when a majority had the incorrect answer.

[16] For example, assume that Candidate A is optimal and B is sub-optimal. Suppose the information on the checklist gives A a total payout of $3.00, and B a payout of $5.00. Assume there is a single error in private information, with a subject writing down "negative" for A's rating when it is actually "positive." Since positive ratings are worth $2.00, and negative are worth -$1.00, correcting this one piece of private information brings A's payout to $6.00, so that the group would have voted for A. Since this is the only error, the counterfactual probability of this group switching to A is .879, the average correction rate for common information mistakes. If instead two mistakes in private information for A need to

check list, with standard errors reported in parentheses below the coefficient values. The coefficient values for both ComCorr and PrivCorr have the expected sign, but neither is statistically significant at conventional levels. The coefficient value for PrivWro is negative and statistically significant ($p < 0.05$), with the estimated effect of a single piece of incorrect private information reducing the likelihood of detecting a HP by 10%. Dropping the PrivWro variable from the regression none of the coefficients is statistically significant at conventional levels.[17]

The group correction factor is inherent to the structure of the HP problem. It has nothing to do with any inadequacies in pooling of information but results from the structure of the HP problem as typically implemented in laboratory experiments, in conjunction with mistakes in recalling information. Quantitatively, the marginal effect of the group correction factor accounts for a little less than half (48.8%) of the failures to uncover HPs in Experiment 2. Further, the fact that mistakes with respect to common information are typically corrected for could well account, by itself, for the oft-noted fact that groups tend to discuss common information more than private information in the caucuses.

Documenting the important role of the group correction factor led us to examine the group decision literature for prior research on the role of mistakes in recalling information in accounting for the failure to discover HPs. As far as we are aware no one else has ever identified or looked at this factor. This raises the obvious question of how relevant are our results to the large number of HP experiments reported in the literature? The answer is: we do not know exactly. However, one criticism that we have encountered is that given our incentive structure, and the abstract nature of subjects' preferences over these characteristics, we might have an exceptionally high error rate. While it is not obvious that this is the case, the closest we can get to this concept given the available data is to calculate the "errors in judgment" present in standard HP experiments and compare it with the mistake rate identified here. For example, in the Stasser and Titus (1985) unshared/consensus treatment the distribution of pre-discussion

---

be corrected to discover the HP, then the probability of this happening is $.879^2$, or .77. If only one of two mistakes needs to be corrected to discover the HP, the probability is $1-(1-.879^2)$, or .99.

[17] PrivWro and PriCorr are negatively correlated, with a correlation coefficient of -0.515 ($p < 0.01$).

information was intended to create a consensus in favor of the sub-optimal choice, B, but resulted in 25% of the subjects preferring the optimal choice, A. If we make the admittedly extreme assumption that the 25% preferring candidate A resulted strictly from "errors in judgment" in evaluating pre-discussion information, and that each characteristic is weighted equally by subjects, then the implied judgmental error rate is 30%.[18] This is well above the 12.7% error rate identified here, which implies that subjects given actual candidate descriptions make more errors in the interpretation and recall of that information than subjects in our experiments which are only given ratings. Although these two error rates are not the same conceptually, they are the best we have to work with.[19]

4. Discussion: One of the most powerful rationales for group decision making is that it provides a mechanism for channeling individual members' knowledge into productive organizational outcomes. The results of an extensive series of experiments reported in the organizational behavior and social psychology literature question this assumption, at least in the presence of a hidden profile problem. In a hidden profile problem each member of a group receives both common and private information about the choice at hand. None of the information is in conflict and the group is presumably working on a common purpose problem where preferences for the best alternative are the same across group members. The disturbing result from these experiments is that freely interacting groups typically fail to choose the best alternative. Reasons for this offered in the literature are that groups (a) spend too much time discussing common information rather than the private information needed to uncover the hidden profile, and (b) disproportionately recall information favoring one's pre-discussion preferred alternative.

---

[18] In this HP experiment each subject had 3 pieces of pre-discussion information supporting the optimal candidate, A, and 8 pieces of information supporting the sub-optimal candidate, B, requiring a 30% error rate in evaluating the 11 pieces of information for 25% of the subjects to prefer candidate A over B.

[19] The reader should not be mistaken in thinking that the judgmental errors bias the group results in favor of candidate A. Rather, this exercise allows us to isolate and compute the errors in judgment present in the standard HP experiment with respect to the characteristics associated with any given candidate.

Social psychological reasons have been offered to account for the group's concentration on common rather than private information. Stewart and Stasser (1995) suggest that when a group member mentions common information, its veracity can be validated by another group member. As a result of such "social validation," this piece of information is more likely to be collectively endorsed by the group and reported in the written protocol of their deliberation. Private information mentioned by the only group member who is aware of it cannot be validated, so it is less likely to make its way into the group's final report. Given the fact that privately held information cannot be validated by any other group member, the group's reticence to discuss or heed it may be seen by the group as entirely reasonable. A second impediment to the expression of private information is that a person might be reluctant to mention a piece of information if it seems to be contrary to the prevailing group opinion (Hartwick, Sheppard, & Davis, 1982). The hidden profile procedure ensures that the pieces of private information do indeed seem inconsistent with the common information, the latter apparently receiving the endorsement from the majority of the group members.

We have reexamined the hidden profile problem using the standard strategy of experimental economists for gaining strong control over subjects' preferences – explicit monetary incentives in conjunction with induced preferences over candidate characteristics. Our motivation for this is that pre-discussion preferences in prior research failed to show the high degree of homogeneity underlying the common purpose assumption. In both experiments our procedures resulted in substantially more homogeneous pre-discussion preferences in favor of the sub-optimal candidate than in the typical hidden profile experiment, with 100% of the groups selecting the optimal candidate with full information at their disposal in Experiment 1. Nevertheless, absent full information, the rate of hidden profile detection was *less* than would be expected with random guessing.

We developed a simple, idealized model that was intended to ensure 100% detection of hidden profiles as a reference point against which to calibrate the effects of known breakdowns in behavior. In terms of the model, a single deficiency—the fact that subjects could only correctly recall 33% of the pre-discussion information—was enough to predict that any one group has less than a 50% chance of detecting the hidden profile in Experiment 1. This result in itself is not novel, as the literature shows that hidden profile detection, although improved with fewer choice characteristics for subjects to remember, is still a major problem (Stasser and Stewart, 1992; Stasser et al., 1995; Fraidin 2004).

We then reduced the information load in Experiment 2 to the point that if the other assumptions of the model were satisfied, 78% of the hidden profiles would have been discovered. However, only 35% were discovered. Our data show that information recall is more heterogeneous than the idealized model assumes and that recall is biased in favor of the preferred candidate based on pre-discussion information. However, each of these has little potential for explaining the shortfall in hidden profile discovery. Instead, we define and find evidence of a "group correction factor" which results from the mistakes inherent in subjects' imperfect information recall. This factor is by far the strongest, as without it the expected rate of hidden profile discovery almost doubles to 66.8%. The reason for this is that given the structure of the hidden profile problem, mistakes with respect to common information favoring a sub-optimal alternative generally get corrected in group discussions, while mistakes in private information favoring the optimal alternative cannot be corrected. Further, the oft-noted tendency for subjects to discuss common information much more often than private information (Stasser and Titus, 2003) is consistent with the group correction factor. To the extent that memory limitations and the group correction factor are significant sources of the failure to discover hidden profiles, they have nothing to do with any deficiencies in how groups process information, interact, or make decisions, but are rather structural elements underlying the presence of a hidden profile.

Even after accounting for the group correction factor, there are still a number of cases where hidden profiles were not detected. One possibility for this shortfall is that group discussion itself is deficient, with individuals unwilling to share information because of the social cost of expressing dissenting information (Schachter, 1951), or perhaps discussion ends after a consensus is formed but before all the relevant information could be reviewed. However, the attention paid to the checklists rules this out, as the decision to vote for a candidate was almost completely dictated by the information on the checklists. Further, the high rate of information on the checklists suggests that subjects shared all the information they could.

Our results have important implications for the managerial/organizational behavior, in spite of the fact that management groups typically convene having at their disposal all the information members wish to bring to a meeting. We would argue that limited recall rates and mistakes in what is recalled in hidden profile experiments serve as proxies for the limited ability of agents to process and express relevant information. Our results suggest that one way to minimize potential hidden profile problems is to ensure some duplication of efforts with respect to identifying candidate characteristics. For example, the devil's advocate (DA) technique (e.g., Cosier, 1978) compels a person or group or persons to critique the process or outcome of a different group. The DA group must therefore be knowledgeable about the inputs to and processes of the group it is supposed to apprise. Thus some redundancy would exist in the DA group and the group they are critiquing, which should in turn increase the probability that the original group's private information would be expressed either by the original group or the DA group questioning the effectiveness of their deliberation. Another way to ameliorate the HP problem would be through process accountability (PA) (Siegel-Jacobs and Yates, 1996). PA focuses on evaluation of the procedure used to arrive at an action. PA is to be distinguished from outcome accountability (OA), under which the outcomes of an action serve as the basis for the evaluation. Siegel-Jacobs and Yates (1996) found that PA encouraged decision makers to take into account more information in arriving at their judgments than they did under OA. If these results generalize to the group situation, one might hypothesize that the private information would be more likely to be expressed and considered under PA than under OA. Wittenbaum and Park (2001) suggest that making group members aware of the domain of expertise of each other member will foster members' willingness both to offer and accept unshared information, on the belief that such information is being offered by someone who is expected to know more in that area. Also, Wittenbaum and Park (2001) suggest that leaders might be trained to solicit and then focus the group's attention on unshared information. Leaders would have the higher status and credibility needed to risk emphasizing such data. However neither the emphasis on PA over OA nor following the suggestions of Wittenbaum and Park (2001) will be able to overcome the principal problem we have identified: Even if unshared information is expressed with greater likelihood, with no one else to verify it, any mistakes in the recollection of such information will go uncorrected. In this respect our results serve to deflate the use of hidden profile experiments as exemplars of pathologies associated with group discussion and information aggregation, as the cause is more closely related to the structure of the problem, than to any deficiencies in how group members interact.

References

Anderson, L. R., C. A. Holt (1997). Information cascades in the laboratory. *American Economic Review* **87** 847-862.

Blinder, A. S., J. Morgan (2005). Are two heads better than one? Monetary policy by committee. *Journal of Money, Credit and Banking* **37** 789-811.

Brodbeck, F. C., R. Kerschreiter, A. Mojzisch, S. Shulz-Hardt. (2007). Group decision making under conditions of distributed knowledge: the information asymmetries model. *Academy of Management Review* **32** 459-479.

Chen, K-Y., L. R. Fine, B. A. Huberman. (2004). Eliminating public knowledge biases in information-aggregation mechanisms. *Management Science,* **50** 983-994.

Cosier, R. A. (1978). The effects of three potential aids for making strategic decisions on prediction accuracy. *Organizational Behavior and Human Performance* **22** 295-306.

Davis, J. H, T. Kameda, M. F. Stasson. (1992). Group risk taking: selected topics. In J. F. Yates (Ed.) *Risk-taking behavior* (pp. 163-181). New York: John Wiley & Sons.

Dellarosa, D., L. E. Bourne, Jr. (1984). Decisions and memory: Differential retrievability of consistent and contradictory information. *Journal of Verbal Learning and Verbal Behavior* **23** 669-682.

Dennis, A.R. (1996). Information exchange and use in group decision making: You can lead a group to information, but you can't make it think. *MIS Quarterly* **20** 433-457.

Fraidin, S.N. (2004). When is one head better than two? Interdependent information in group decision making. *Organizational Behavior and Human Decision Processes* **93(2)** 102-113.

Goeree, J. K., T. R. Palfrey, B. W. Rodgers, R. D. McKelvey. (2007). Self-correcting information cascades. *Review of Economic Studies* **74** 733-762.

Harrison, D. A., K. J. Klein. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review,* **32** 1199-1228.

Hartwick, J., B. H. Sheppard, J. H. Davis. (1982). Group remembering: Research and implications. In R. A. Guzzo (Ed.), *Improving group decision making in organizations* (pp. 41-72). New York: Academic Press.

Hollingshead, A. B. (1996). The rank order effect in group decision making. *Organizational Behavior and Human Decision Processes* **68** 181-193.

Klein, K. J., D. A. Harrison. (2007). On the diversity of diversity: tidy logic, messier realities. *Academy of Management Perspectives* **21** 26-33.

Larson, J. R., Jr., P. G. Foster-Fishman, C. B. Keys. (1994). Discussion of shared and unshared information in decision-making groups. *Journal of Personality and Social Psychology* **67** 446–461.

Rulke, D. L., J. Galaskiewicz. (2000). Distribution of knowledge, group network structure, and group performance. *Management Science* **46** 612-625.

Schachter, S. (1951). Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology* **46** 190-207.

Schulz-Hardt, S., D. Frey, C. Luthgens, S. Moscovici. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology* **78** 655-669.

Siegel-Jacobs, K, J. F. Yates. 1996. Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes* **65** 1-17.

Stasser, G., D. Stewart. (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., L. A. Taylor. (1991). Speaking turns in face-to-face discussions. *Journal of Personality and Social Psychology* **60** 675–684.

Stasser, G., W. Titus. (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., W. Titus. (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.

Stasser, G., W. Titus. (2003) Hidden profiles: A brief history. *Psychological Inquiry, 14*, 304-313.

Stasser, G. (1988) Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology* **24** 393-422.

Stasser, G., D. D. Stewart, G. M. Wittenbaum. (1995). Expert roles and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology* **31** 244-265.

Stewart, D. D., G. Stasser. (1995). Expert role assignment and information sampling during collective recall and decision making. *Journal of Personality and Social Psychology* **69** 619-628.

Sunder, S. (1995). Experimental asset markets: a survey. In J. H. Kagel and A. E. Roth (Eds.) *The handbook of experimental economics* (pp. 445- 500). Princeton: Princeton University Press.

Thomas-Hunt, M.C., T. Y. Ogden, M. A. Neale. (2003). Who's really sharing? Effects of social and expert status on knowledge exchange within groups. *Management Science* **49(4)** 464-477.

Weber, S. S., L. M. Donahue. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: a meta-analysis. *Journal of Management* **27** 141-162.

Wittenbaum, G. M., E. S. Park. 2001. The collective preference for shared information. *Current Directions in Psychological Science* **10** 70-73.