Information Sharing for Collective Sensemaking

Yuqing Tang*, Christian Lebiere*, Katia Sycara*, Don Morrison *, Michael Lewis†, and Paul Smart‡
*Carnegie Mellon University, Pittsburgh, US
†University of Pittsburgh, Pittsburgh, US
‡University of Southampton, Southampton, UK
Email: *{yuqing.tang,katia}@cs.cmu.edu, {cl,dfm2}@cmu.edu, †ml@sis.pitt.edu, ‡ps02v@ecs.soton.ac.uk

***Abstract*—Group decision tasks that require pooling of information to reach the best decision have been studied across a variety of disciplines over the past thirty years. The crucial question of what makes these tasks so difficult, however remains unanswered. Various hypotheses include inefficiency in sharing information leading to decisions based on incomplete information or cognitive inefficiencies in processing and storing information arriving in a piecemeal fashion. The present study attacks this problem from two directions. Human experiments are used to compare decisions between groups manipulated to receive and share information in raw and aggregated forms and mixed groups comprised of humans and software agents. To shed light on cognitive limitations that may affect performance, an ACT-R cognitive model of group members was constructed and its results compared to human data.**

1. Introduction

In many group decision tasks, group members may possess unique information which, if completely shared and used for decision making, would help the group arrive at optimal decisions. Possessing unique information by group members has been called the *hidden profile* paradigm and was first introduced by `strasser1985. Since the introduction of the hidden profile paradigm, numerous studies have been conducted to study various aspects of information sharing and pooling, discussion of common vs unique information, group performance, etc. The results of these studies have not all been in agreement (see [6] for a recent meta-analysis of 60 papers on the hidden profile group decision making). Prior studies addressed both the process (how information is shared) and product (correctness of selected alternative). The dominant finding about process has been that groups focus discussion on common rather than unique material. In fact, groups spend much more time discussing common rather than unique information [6]. This suboptimality is reflected in group decisions where hidden profile groups were consistently much less likely to select the correct alternative than groups with full information. Related findings, however, suggest that coverage (the degree to which relevant information has been shared) may be more important than discussion in improving decisions.

A key consideration missing from this prior research is the source of the large difference in performance between groups receiving common information and those who must tease it out through discussion and questioning. Another limitation of prior research is that almost all existing studies report only counts of information items pooled without regard for the content of the information shared. Additionally, counting ignores differences in the usefulness of the information pooled. Certain pieces of information may be more useful for the task solution than others. Finally, current research cannot distinguish between the effects of incomplete coverage and those associated with cognitive processing limitations in forming, modifying, and rejecting hypotheses as information is progressively revealed.

Our paper addresses these literature limitations. We have designed a study that compares the performance of human teams voluntarily sharing information under conditions comparable to those in reported hidden profile studies with artificial teams in which a single human teammate is provided full information by *Bots* posing as human teammates. Comparison between these two conditions can elucidate the extent to which coverage can account for differences in performance between hidden profile and full information groups, because the *Bots*, by design, supply all relevant information while human team mates choose what information to pass on in the manner of hidden profile experiments. Although a true control providing human teams with full information at the outset is not available, the size of the effect between human and artificial teams can provide indications of the degree to which information sharing alone could account for the large differences commonly observed between hidden profile and full information teams.

Our study also addresses the prior limitations of the literature as to the contributions of coverage and cognitive limitations to the hidden profile deficit. In particular, we consider a *guess* and *no-guess* condition. In the *guess* conditions, decision makers share their best "guess" of the answer to their role's designated question. This *guess* is visible to all the other group members. It is hypothesized that because a *guess* represents a summary information of the deliberation of another team mate, the decision maker does not need to spend cognitive effort storing data, reviewing, and revising hypotheses.

To deal with limitations of prior research in considering cognitive processing characteristics in information sharing and hypothesis formation and modification, we have developed a human computational model based on the ACT-R cognitive architecture [1] for performing the task to help identify memory, attentional and other mechanisms that may be implicated in the task performance. This model's mechanisms, such as spreading activation and reliance on probability estimation rather than logic in judging data and hypotheses, replicate well known cognitive biases such as availability, recency, and confirmation bias [5].

We present the design of an experiment intended to address these issues, specifically by allowing decision-makers to selectively share detailed information or a high-level guess regarding a variety of questions that can be answered using the information. We subsequently present an analysis of the experimental results as well as results comparing the ACT-R model to the human data. We present conclusions and plans for future work.

2. Experimental Task

The task used simulates, using textual information, an artificial world region with political unrest. It requires four cooperating subjects to discover a variety of details and draw conclusions regarding an impending terrorist attack. The data underlying this task are from ELICIT [2]. Facts[1] from which these conclusions can be drawn are released to the subjects in 3 stages over time.

Each of the four subjects is given a role, i.e. to answer a different question about the attack: who, where, what and when. In other words, in each trial there are four roles (who, what, where, when) each assumed by a different subject. Role assignment is done by the system at the beginning of each trial. There are four different trials, the first of which is a training trial, followed by *three experimental trials*, each with a different set of facts concernign different groups, targets, and country names. The answer to the who *question* is the name of the group expected to conduct the attack; group names are denoted by colors, such as the "gold group" or the "violet group." The answer to the where question is a country name; country names are denoted by Greek letter names, such as "Chiland" or "Omegaland." The answer to the what *question* is the intended target, such as "embassy" or "military base." The answer to the when question has a four-fold structure, consisting of month name, day of the month, hour on a twelve hour clock, and "AM" or "PM." While not the subject of any of the questions, there are also individuals, who serve as links connecting some of the facts presented to subjects. Individuals are named after animals, such as "the Lion" or "the Jackal."

The facts delivered to the subjects are sentences. Some are simple and immediately useful, such as "The attack will be at 11:00." Others are more complex, and must be combined with other information to be useful; for example, "The Azure and Brown groups prefer to attack at night," or "The Lion is known to work only with the Azure, Brown, or Violet groups." Some of the facts delivered are essential for constructing correct answers, others are helpful but not essential, and still others are mere noise, contributing nothing to correct answers.

In each wave, the system gives a set of facts to each participant/role. These sets of facts are disjoint, in other words no fact in a set is given to more than one subject. A participant may receive some facts relevant to determining the solution to its role and other facts that may not be useful to her but possibly useful to other roles. A participant is free to determine whether and to whom to forward facts she has received. For example, the participant in the role of what may receive the fact "The Violet group prefers to work in Alphaland or Omegaland". This fact is relevant to the who role and the where role although it is not directly relevant to the what role (though later on it may be indirectly relevant if information that links the Violet group to some specific target becomes available).

For each trial, 68 facts are distributed in three waves. Each wave contains roughly 1/3 of the 68 facts. Between two consecutive waves, the subjects have 5 minutes to process a wave of new facts. Each trial lasted 15 minutes. At the end of each trial, the subjects submitted their best answer to the *question* of their role.

The experiment had the following conditions: First condition was the "only-humans; *no-guess*" condition where the subjects received the facts in 3 waves, made decisions on information sharing, shared information and made their final answer within the 15 minute trial interval. Note that the subjects were not encouraged to give their answer as early as possible but were left to decide to give their answer as soon as they felt confident about it. The second condition was the "only-humans; *guess*" condition which is same as the "only-humans; *no-guess*" condition except that each subject made a guess as to the answer to her *question*, along with her confidence in the *guess* on a five point scale, after each wave. All *guesses* were shown to the other agents. The third condition was the "robot-humans;*no-guess*" where a human was teamed with three software agents. However, since the communication was via a computer interface, the human did not know that her team mates were computer programs. The agents always forwarded all the pieces of information that were useful to each role. The fourth condition was "robot-humans;*guess*" where the human was teamed with software agents and where she made guesses at the end of each wave. The agents also publicized their *guesses*. The agents were programmed to make the correct guess after the first wave and their *guess* did not change after the second or third. By contrast, in the human only condition, the *guess* of a person could change as he accumulated more information in each subsequent wave.

The four subjects interacted with the system and with each other through a web-based user interface, Figure 1, implemented with HTML and JavaScript. This interface is divided into several panes. One, on the right, summarizes the player's current role (who, where, what or when), describes the names and roles of the other players, and allows access to the instructions for reference.

The most prominent pane of the interface is the inbox, to which new facts are delivered. These may be new facts, delivered by the system; or they may be facts forwarded by another subject. Facts are normally displayed here in a partially obscured form, with only a few keywords, such as "Yellow," "Magenta" and "Green," legible, the rest of the text being replaced with ellipses. The user can click on a fact to cause the full text to be presented. When the mouse pointer is moved off the fact, it is partially obscured again (see Figure

[1]ELICIT calls these statements "factoids"

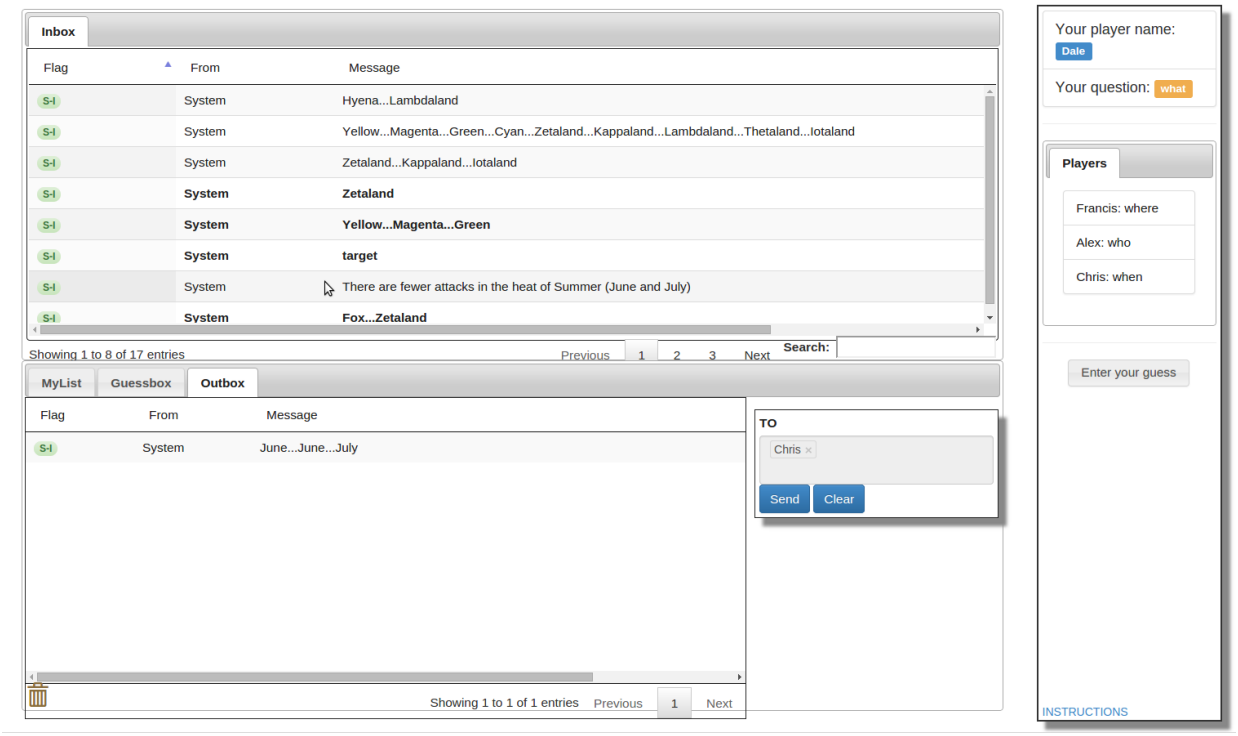


Figure 1: The user interface of the experiment

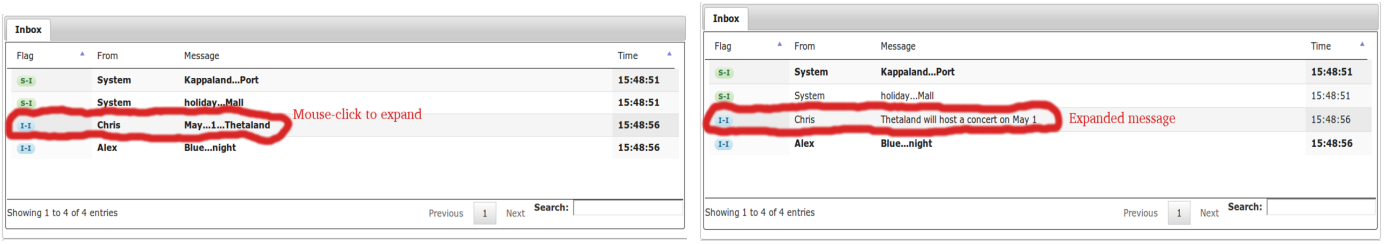
2). The aim of this design was to enable us to get a handle on which facts each user was *paying attention* to and for how long. This information is useful to determine which facts they potentially considered most useful and also determine memory limitations.

Below the inbox is a pane multiplexed for three purposes: outbox, mylist and guessbox. When used as the outbox facts can be dragged to it, and forwarded to other subjects, in whose inbox they will appear. When used as mylist, facts can be dragged to it for future reference. While users can use this for whatever purpose they choose, it is expected that those who do employ it will use it to consolidate facts they suspect are important for answering their own *question*. Facts in mylist, as in the inbox, are normally partially obscured, and must be clicked to be read in full.

Although ELICIT at first glance differs from the classical hidden profile task along a number of dimensions, on closer examination the essential elements of these tasks are the same. The primary differences are: 1) In hidden profile studies group members share information and discuss it in order to arrive at a group decision. In ELICIT, group members share information in order to arrive at individual decisions. We believe that this distinction is less telling than it appears because hidden profile experiments are designed such that "the common information favors a suboptimal decision alternative, whereas all the unique information combined reveals the optimal alternative" [7]. Because the optimal decision is predetermined when all the information is pooled, the process being investigated in the hidden profile experiments is information sharing and not decision making per se. Decision makers in ELICIT face the same situation in that their decisions rely on information shared by the group and are largely predetermined by what is shared. 2) In hidden profile studies group members are provided with both common and unique information. Much analysis has involved the attention paid to common and unique information in group discussions. In ELICIT all information is unique until it has been shared. This prevents the corresponding comparisons between the use of common and unique information. Measures of coverage and sharing of relevant information, however, are available and allow comparisons. The cognitive limitations hypothesis is consistent with the hidden profile observation [6] that groups focus on common (prior) information and provides an alternate explanation, namely that attentional and memory limitations prevent the consideration of the entire set of information when making decisions. Rather, a small subset of information is favored, often that which is more directly available (availability bias), has been accessed more recently and is thus more active in memory (recency bias), or has been the lead hypothesis and thus has been repeatedly reinforced (confirmation bias). Data on participants' attention paid to different pieces of information are available and analyzed from our ELICIT experiment.

3. ACT-R Cognitive Model

The ACT-R model uses the ACT-R cognitive architecture [1] and in particular leverages the activation calculus in declarative memory. This modeling approach reflects the fact that performance in this task heavily relies on retrieval of information from memory, and the activation processes in



(a) Factoid in keywords

(b) Factoid content expanded

Figure 2: Experimental UI on tracking subject attention

ACT-R declarative memory provide powerful mechanisms to guide information retrieval as well as embody limitations on the storage and retrieval of memories under demanding conditions.

The task is decomposed into three component sub-tasks: information sharing, inductive inference, and probability estimation. The goal of the information sharing subtask is to determine which facts to share, and if so with which of their teammate(s). The basic approach is to share facts that are semantically related to the *question* domain of the teammate, e.g., share facts containing location information with the person in charge of the 'who' *question*. The implementation leverages the ACT-R partial matching mechanism that retrieves chunks in declarative memory by combining activation with semantic similarities.

The goal of the inductive inference subtask is to determine the relevance and applicability of various facts to the specific *question*. The approach is to activate facts whose context is associated with a specific *guess* or answer. The implementation leverages the ACT-R spreading activation and base-level learning mechanisms. Finally, the goal of the probability estimation task is to determine the probability of each candidate answer given various facts and their activation. This approach is grounded in the assumption to represent each fact as a rough probability estimate over the given options. The probability estimation process leverages the blending memory retrieval mechanism to generate aggregate estimates.

The information sharing model assumes that people use primarily simple heuristics in determining whether information is relevant to another decision maker. In this case, the assumption is that subjects share a fact if they contain information of the same semantic domain as that person's *question* (e.g., location information if the given *question* is 'who'). While this approach is well short of optimal it has the advantage of being highly efficient and avoids assuming knowledge of the other decision makers processes that is unlikely to be available. To avoid implementing this approach using large numbers of ad hoc heuristic rules (e.g., one for each combination of question and answer), pattern matching processes in memory are used instead. Each fact is encoded as a set of semantic keywords reflecting the key information contained in the sentence. For each subject and given *question*, each keyword is matched against the various *questions* to determine if it contains information related to that *question's* domain. The ACT-R partial matching mechanism is used by setting high semantic similarities between concepts of a common domain (e.g., locations like 'psi' and 'chi', and a *question* like 'where'). This will result in the relevant *question*(s) retrieved for each fact, indicating the relevant decision maker with whom to share that fact. The mismatch penalty scaling factor controlling the partial matching process will determine the selectivity of the process. Thus varying the scaling factor can determine the overall willingness to share information, leading to individual differences reflected in more conservative or widespread sharing.

The inductive inference model is used to reflect dependencies between answers to different *questions*. For instance, an answer (or guess) of a given location to a 'where' *question* (e.g., 'psi') will raise the relevance of facts mentioning that location when answering other *questions* (e.g., 'who'). The most natural way to implement that dependency process is to use the spreading activation mechanism. To do that, each fact is associated with a context element representing its dependency upon another answer (guess). When a given answer is processed, related facts are retrieved from memory by spreading activation from that answer to the facts including that answer as context. Those facts receive a boost in activation from the base-level learning process, making them more salient in the subsequent probability estimation process. This boost can result in well-known cognitive biases such as availability bias. This approach is similar to the model of the impact of memory availability in the model of sequential diagnostic reasoning of Melhorn et al (2011).

The probability estimation model follows the instance-based learning (IBL) modeling methodology [3]. To provide for finer discrimination in judgment and ensure the ability to gradually accumulate evidence from a stream of individual facts, the basic problem of determining the most likely candidate answer for each *question* is formulated as a goal to assign a probability to each potential answer. The goal is defined as a chunk of type hypothesis that contains three slots:

- Question: the representation of the *question*, i.e., who, what, where and when
- Answer: the representation of each possible answer, e.g., various groups for who
- Probability: a probability value assigned to the question-

answer pair

This representation follows the general IBL pattern of context (question), decision (answer) and outcome (probability). In keeping with the instance-based methodology, this representation is used both for facts as well as goals. Specifically, most facts are transformed into chunks of this type if they make a strong assertion about a given question. For instance, if the fact rules out a particular group, a hypothesis chunk will be created (or reinforced if it already exists) stating (who, group, 0). Conversely, if it strongly implies a group's involvement, the chunk (who, group, 100) will be created. If the fact mentions the possible involvement of n groups, then a separate hypothesis chunk is created for each group with a probability of $1/n$, reflecting mutually exclusive participation.

Of course, those assertions are not literally correct–rather the intent is to provide the basis for a rough estimate of relative probabilities based on the information provided. More precise facts (e.g., stating actual probabilities, or using qualifiers such as likely or probably) could be used to create more accurate chunk encodings. When the model is asked to generate a *guess* to a *question*, it iterates through all the possible answers (e.g., all the groups for a who *question*) and generates a probability estimate for each using the blending mechanism used for memory retrievals [4]. During memory retrievals, each chunk in memory has an activation that reflects factors such as recency, frequency, and degree of match to the requested pattern. Recency is factored through a power law decay from the time that the chunk is created. Frequency reflects a power law of practice of the numbers of times that a chunk is strengthened following rehearsals. For degree of match, we assume for simplicity that each question and answer are distinct and no similarities are defined. Blending retrieval then assigns for a given question-answer pair a probability to each chunk matching that request (in general, there will be several) reflecting a softmax (Boltzmann) distribution of chunk activations given a certain amount of noise. Those probability estimates for each chunk associated with the question-answer pair are then blended according to a weighted average of the chunk probabilities (assuming linear similarities over the probability space [5]). The probability estimates are not normalized but instead the largest one is selected to generate the guess. All parameters controlling the behavior of the model are left at their default values: the base-level decay rate is 0.5, the mismatch penalty is 2.5, the activation noise is 0.25, and the blending temperature is 0.4.

Note that, as mandated by the ACT-R theory, the hypothesis goals generated to provide the guess become themselves chunks in memory, as are guesses received from other agents. This can give rise to cognitive biases such as confirmation bias, where a strong initial estimate leads to overoptimistic estimates later despite contradictory evidence.

4. Results

Sixty subjects, divided into twelve groups of five, were recruited and finished the task. While they did not know how they were divided, four of each five worked cooperatively together, and the fifth worked separately, with three *bots*.

Table I shows the logistic regression of correct decisions on *bot* (group type Bot vs. human), *guess* (with or without *guess* on other *questions*), and *question* (who, what, or where). Main effects were found for both Bot and *question* along with a 2-way interaction for *guess* x *question* and a 3-way interaction among all terms.

Table I

Logistic Regression on Correct Decisions				
Coefficients	Estimate	Std. Error	z Value	P($>\|z\|$)
(Intercept)	$-.06908$	0.4664	-1.481	0.13860
robot	3.2338	1.4901	2.170	0.03000
guess	1.9183	0.6947	2.761	0.00576
question	0.2481	0.3541	0.701	0.48352
robot x guess	-2.8053	2.0008	-1.402	0.16089
robot x question	-1.6171	0.9668	-1.673	0.09439
robot x guess x question	2.9267	1.4176	2.064	0.03897

A. Information Sharing

We have hypothesized three factors implicated in producing the large difference in accuracy in hidden profile experiments [6] observed between groups provided with full information at the outset, called the *manifest profile*, and those required to share information over the course of a discussion.

- deficiency in sharing of information (do players share the right facts)/coverage of information (degree to which relevant information has been shared),
- cognitive processing limitations (deficits due to sequential processing of information)
- attention to information (degree to which relevant information received from others has been accessed)

Decisions were scored for correctness (correct/incorrect) and effects of conditions tested using logistic regression.

Deficiency in sharing information/coverage Deficiency in sharing information leading to incomplete coverage of what is known within the group is likely to be one of the contributors to the underperformance of hidden profile groups. This deficiency is presumed to arise because group members do not recognize the usefulness of their unique information to other members of the group. In our experiment, the effects of deficiency in voluntary sharing of information can be examined by comparing the performance (correct decisions) of the Bot groups in which the human decision maker received all information relevant to his decision with human groups in which only information thought relevant by other players was shared. The extent to which these performances vary indicates the degree to which incomplete information rather than limitations in information processing led to incorrect decisions. The logistic regression contrasting human and Bot conditions found superior performance for decisions made in the Bot condition. The accuracy of decision in the human condition was 43%, whereas in the Bot condition, it was 72% ($z = 2.170$, $p = .03$). This advantage held for each of the decision types,

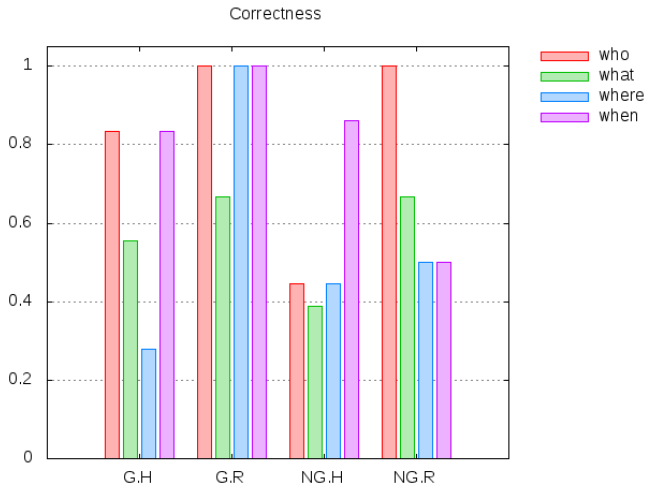


Figure 3: Correctness of results across all *questions* and and the *guess* and *no-guess* conditions. G=guess, NG= no guess; H=human, R=robot. Therefore GH denotes the condition of guess for human only teams and NGR denotes the conditin of no-guess in the robot human team.

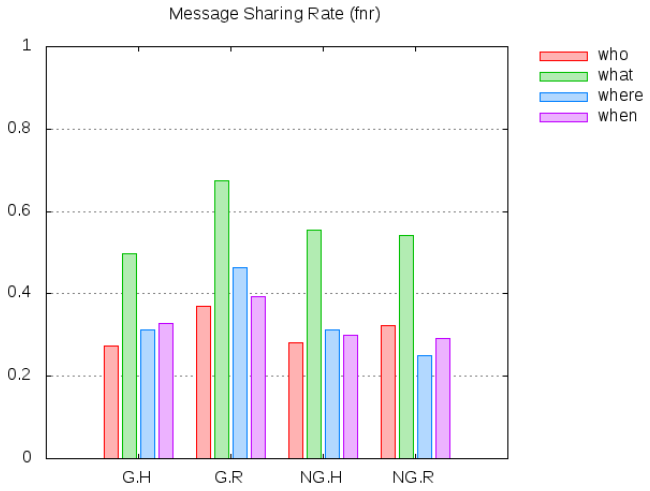


Figure 4: Rate of false negatives, i.e. messages that should have been shared but were not

namely: what, where accuracy in condition human= 40%, and in condition bot= 67% ($z = 2.021$, $p = .043$); who, accuracy of condition human= 54%, and condition bot= 80% ($z = 2.385$, $p = .017$), and where, accuracy in condition human= 33%, and in condition bot= 58% ($z = 3.719$, $p < .001$).

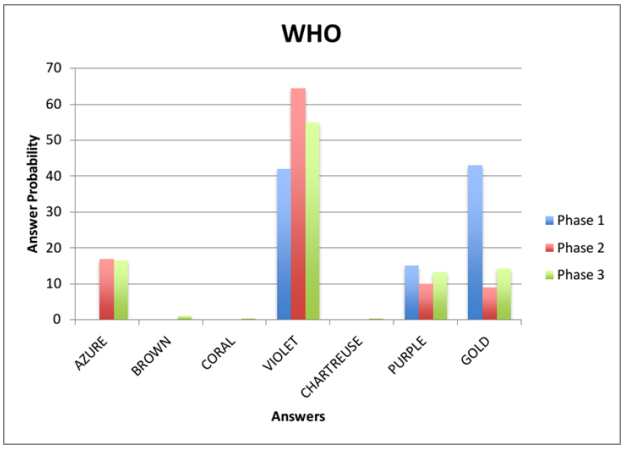
Another element that is important and has not been examined in the literature is the content of the information shared. In the ELICIT experiment, the facts that were relevant to a particular decision were notated in the data, however this relevance was not known to the participants. Since we had this "ground truth" we were able to determine the rates at which different categories of facts were shared. The categories we examined were:

- TP: true positive = the message should be sent and it is sent
- FP: false positive = the message should not be sent but it is sent
- TN: true negative = the message should not be sent and it is not sent
- FN: false negative = the message should be sent but it is not sent
- TPR (true positive rate) = TP / (TP + FN)
- FNR (false negative rate) = FN / (TP + FN) = 1 - TPR
- FPR (false positive rate) = FP / (FP + TN)
- TNR (true negative rate) = TN / (TN + FP)

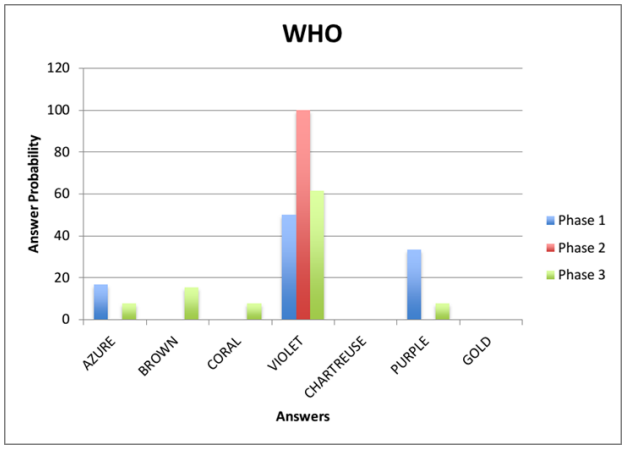
Figure 4 shows the rate of sharing of false negative messages. From the figure we see that for example, for the what *question*, the rate of false negatives was higher than others. This correlates with the low correctness for the what *question* (see figure 3).

Cognitive Processing Limitations The cognitive processing limitation explanation for hidden profile underperformance, was also supported by these data. The *guess* manipulation in the experiment provided decision makers receiving *guesses* with preprocessed/summary data, freeing them from the obligation to acquire, store, and revise hypotheses about answers to the other *question*s needed to support their own decisions. As a consequence, if deficits in performance were occurring due to difficulties in remembering facts and revising hypotheses, these difficulties should be reduced by the *guesses* provided to the decision makers. This manipulation, however, does not eliminate the need to acquire and store information and to maintain and revise hypotheses about the decision maker's own *question* so the reduction in cognitive load is only partial. As the regression shows, providing *guesses* led to marked improvement overall ($z = 2.76$, $p < .006$). As indicated by the 3-way interaction, however, support through guessing was particularly effective in the Bot condition where correct information was always provided.

Attention to Information Our data do not allow direct comparison with the hidden profile finding [6] that "common" information receives more attention in discussion than information that is unique to individual decision makers. We do, however, have detailed records through the content expansion mechanism of the information to which users have attended (see Figures 5 and 6). Analyzing these data we find that the percentage of relevant information among that which is attended is related to improved performance ($z = 1.92$, $p < .05$).

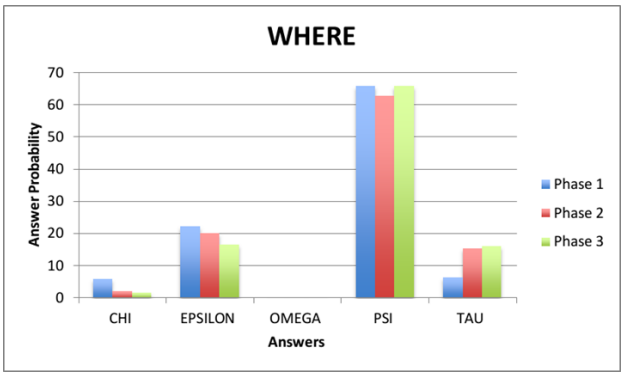


(a) ACT-R

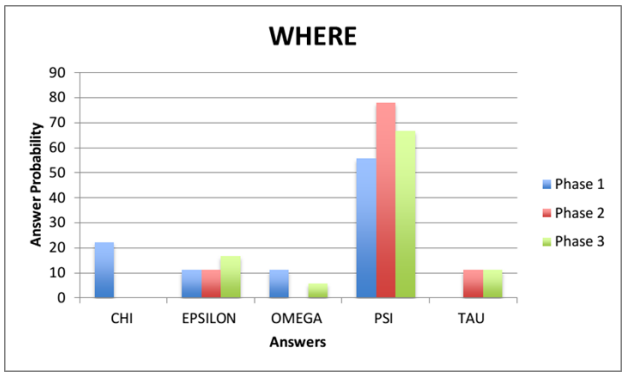


(b) Human

Figure 7: Results of ACT-R and Human on "who" question



(a) ACT-R



(b) Human

Figure 8: Results of ACT-R and Human on "where" question

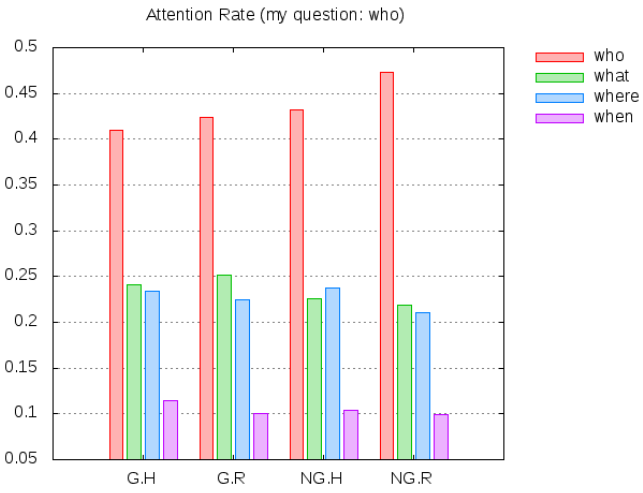


Figure 5: Attention of players assigned "who" question

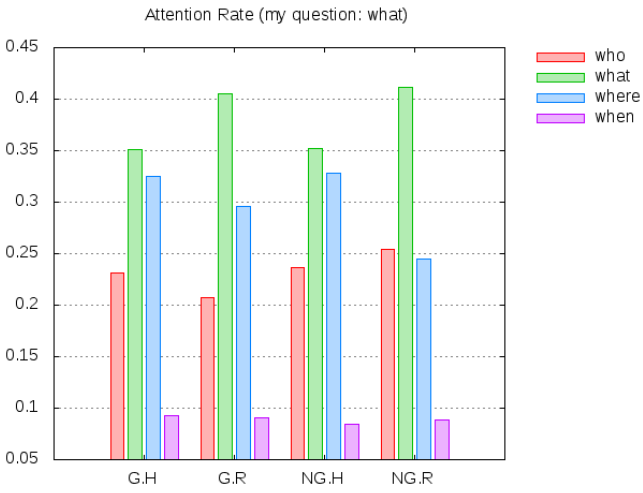


Figure 6: Attention of players assigned "what" question

B. Comparison of ACT-R model and humans

Among the 60 subjects who participated in our experiments, 15 of them (including the subjects who worked with *Bots*) answered the "who" *question* for the fact set "1aGMU17", one of the three fact sets in the experimental trials. Sample results for the "who" *question* are presented in Figure 7 of fact set "1aGMU17". Probability estimates for each possible answer (i.e., all groups) are presented for each of three waves of facts. Among all the human participants, 50% reached to the correct answer, "the Violet group", after seeing the first wave of facts. After seeing the second wave of facts, 100% of the participants reached the correct answer. However, after the third wave, about 40% of the participants were confused by the new facts and changed their answers from the correct one. In other words, for the human subjects, the initial estimate for the violet group (the correct answer, as it turns out) is the highest but closely matched by competing groups such as purple and gold (purple and azure for human subjects). This reflects the lack of complete information to conclusively decide between competing alternatives after the first batch of facts, leading the probability estimation process to fragment its estimates. Following the second batch of facts, violet emerges as the strong consensus answer (unanimous in the case of human subjects) following confirming evidence. However, the estimate for the violet group falls after the third batch of facts due to a dilution effect from a number of facts mentioning other possibilities. Those other three or four possibilities do not rise to seriously challenge the leading answer, but they introduce enough doubt to drag down its probability. The fact that an answer (correct as it turns out) that was (near) unanimous is thrown in doubt by incidental evidence that does not in any way invalidate it establishes the true nature of the process as one of gradual, approximate, implicit aggregation rather than precise, symbolic, flawless deduction. Note that these results were generated without reflecting the effect of previous guesses on later phases. This would be a case where confirmation bias could actually lead to a correct final answer by strengthening the correct guess based on the effect of early evidence and prevent its reconsideration in light of additional facts. As is generally the case, bias is always relative to its environment.

The dynamics of results for the "where" *question* are substantially different (see Figure 8). There, the leading (and correct) hypothesis ("Psi") quickly establishes itself. While no alternative establishes itself as a serious competitor, they do not disappear either but rather sightly fluctuate with each new wave of fact incapable of decisively tilting the balance in either direction. This reflects the cognitive and process limitations of model and human subjects that, despite the potential availability of all needed information, failed to conclusively reach the correct answer.

While both data analysis and model development are preliminary, they highlight interesting emerging effects. In both the human data and the cognitive model, rather than following a linear path, the deductive processes faced with a constant stream of facts induce a fluctuation in beliefs that reflect a potentially rich dynamic. This evidence for information accretion in both human subjects and the ACT-R model support the cognitive limitation explanation for the poor performance of hidden profile groups.

5. Discussion

The ELICIT intelligence analysis simulation has provided an opportunity to investigate well documented information sharing effects found in hidden profile studies. By eliciting multiple decisions per experiment, direct attentional measures through mouse-overs, and informationally optimal groups through use of *bots* our experiments have allowed us to entertain questions which could not easily be addressed within the conventional paradigm. One difference lies in the magnitude of effects found within the two experimental settings. A meta-analysis of 65 hidden profile studies [6] found an eight fold difference (positive odds ratio of 8.05) between manifest profile and hidden profile groups. Differences observed in our experiments were more modest. Our ELICIT experiments do not have a full equivalent of a manifest profile control but do have a *Bot-Guess* condition in which decision makers received complete information with a substantial amount already aggregated and abstracted. Under these conditions a difference of approximately 2:1 in correct decisions was observed suggesting that the information processing load imposed by integrating arriving information is likely a significant, though not sole, contributor to the hidden profile deficit.

Coverage, the extent to which relevant knowledge was shared with the group was found in [6] to be the primary determinant of hidden profile decision quality. Our experiment found a similar effect with the Bot conditions in which decision makers received full information outperforming human groups where information sharing was by choice. The inclusion of *Bots* within human decision making studies provides a methodological advance for untangling the effects of cognitive limitations due to incremental information presentation and the extent information coverage. *Bot* conditions allow precise manipulation of information sharing which cannot be obtained in all human experiments where coverage must be assessed by monitoring information brought up in discussion. By allowing the experimenter to ensure full information sharing while supplying information incrementally, our *Bot* methodology could allow direct comparison between manifest and hidden profile decision making while controlling for coverage.

Although our current data analysis does not distinguish between attention paid to information forwarded to a decision maker by other players (corresponding to the sharing of unique information in hidden profile), from information provided directly in the three waves (corresponding to common information that could be processed as a unit), our information tracking and attention measuring mechanisms would allow such labeling so that the relations between coverage and attention found in hidden profile studies could be addressed in the ELICIT context.

The primary contribution of this work has been to advance the role of cognitive limitations due to serial processing of shared information to decision making deficits in hidden profile studies and other situations requiring the pooling of information. The *Bot* methodology we have introduced allows control of information coverage and presentation in ways otherwise infeasible. The strong similarities between the ACT-R model which incorporates cognitive mechanisms producing deficits for sequential information presentation and human performance at the ELICIT task further supports this explanation.

6. ACKNOWLEDGMENT

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defense and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defense or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

[1] J. Anderson and C. J. Lebiere. *The Atomic Components of Thought*. Erlbaum, Mahwah, N.J., 1998.

[2] Kevin Chan and S Adali. An agent based model for trust and information sharing in networked systems. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2012 IEEE International Multi-Disciplinary Conference on*, pages 88–95. IEEE, 2012.

[3] C. Gonzalez, J. F. Lerch, and C. Lebiere. Instance-based learning in dynamic decision making. *Cognitive Science*, 27:591–635, 2003.

[4] Christian Lebiere. The dynamics of cognitive arithmetic. *Kognitionswissenschaft Special issue on cognitive modelling and cognitive architectures, D. Wallach and H. A. Simon (eds.)*, 8:5–19, 1999.

[5] Christian Lebiere, P. Pirolli, R. Thomson, J. Paik, M. Rutledge-Taylor, J. Staszewski, and J. R. Anderson. A functional model of sensemaking in a neurocognitive architecture. *Computational Intelligence and Neuroscience*, 2013.

[6] Li Lu, Y Connie Yuan, and Poppy Lauretta McLeod. Twenty-five years of hidden profiles in group decision making a meta-analysis. *Personality and Social Psychology Review*, 16(1):54–75, 2011.

[7] G. Strasser and W. Titus. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(1):1476–1478, 1985.