Pitfalls of Information Access with Visualizations in Remote Collaborative Analysis

Aruna D. Balakrishnan, Sara Kiesler, Aniket Kittur
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15232 USA
{aruna, kiesler, nkittur}@cs.cmu.edu

Susan R. Fussell
Department of Communication
Cornell University
332 Kennedy Hall
Ithaca, NY 14853
sfussell@cornell.edu

ABSTRACT
In a world of widespread information access, information can overwhelm collaborators, even with visualizations to help. We extend prior work to study the effect of shared information on collaboration. We analyzed the success and discussion process of remote pairs trying to identify a serial killer in multiple crime cases. Each partner had half of the evidence, or each partner had all the available evidence. Pairs also used one of three tools: spreadsheet only (control condition), unshared visualizations, or shared visualization. Visualizations improved analysis over the control condition but this improvement depended on how much evidence each partner had. When each partner possessed all the evidence with visualizations, discussion flagged and pairs showed evidence of more confirmation bias. They discussed fewer hypotheses and persisted on the wrong hypothesis. We discuss the possible reasons for this phenomenon and implications for design of remote collaboration systems to incorporate awareness of intermediate processes important to collaborative success.

Author Keywords
Experiment, information sharing, information visualization, information overload, confirmation bias, empirical studies, computer-mediated communication.

ACM Classification Keywords
H5.3. Group and Organization Interfaces [Computer-supported cooperative work].

General Terms
Human Factors.

INTRODUCTION
As last year's meltdown of the financial markets made clear, the world has become significantly interconnected. In many domains of analysis—science, business, criminology, epidemiology, government, and intelligence—the amount of data that must be collected, perused, and analyzed to solve problems is huge. One result of massive information collation and sharing is that the sheer body of information can exceed the unaided capacity of individual analysts.

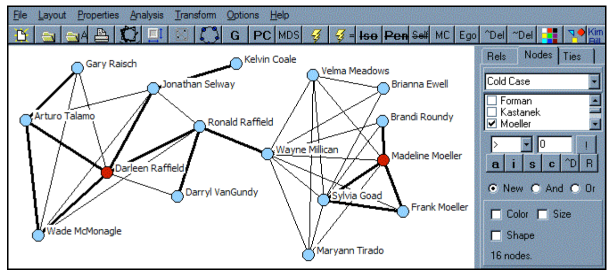

Figure 1. Visualization of case evidence in NetDraw, a network diagram tool [1].

To make headway on ballooning information, two main approaches have been taken. *Social* solutions make analysis a collaborative process. A long-held vision in CSCW is to improve distributed access to data for collaboration [12]. By having full information access and considering the evidence together, collaborators can derive better conclusions from the data. In short, two heads are better than one. *Cognitive* solutions, in contrast, focus on enhancing individuals' cognitive capacity through visualization and other analytical tools that help people process more data, more rapidly [e.g., 3].

Collaborative analysis combined with visualization tools might be an ideal solution to the information overload problem. Visualizations as in Figure 1 have been shown to facilitate collaborative analysis [1,9,26,27]. At the same time, visualizations can fail to overcome coordination costs that arise from the time spent, and possibly wasted, in discussion [34]. Cognitive biases, particularly confirmation bias—the tendency to seek out information that confirms what one already thinks, and avoid information that disconfirms it—can cause analysts to persist on the wrong hypothesis [31]. In preliminary studies, we asked participants to work on an analysis task either individually or with a partner through Instant Messaging (IM). A network visualization tool improved analysis overall, but collaborative analysis was less successful than individual

CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.

analysis. This result suggests that we need to learn more about the process of analysis when collaborators are using visualization tools so that we can improve these tools to overcome coordination costs and cognitive biases.

In the current paper, we examine remote analyst pairs collaborating on the serial killer task described in Balakrishnan et al. [1]. Success on this task depends on insight when combing through hundreds of pieces of evidence. We examine how the distribution of evidence (each partner has all the evidence or each has half of it) and the availability of visualization tools change how the pairs discuss the evidence and their problem solving success.

Distribution of Evidence

For organizational, legal, political, and other reasons, collaborators may have different access to the myriad of raw data or evidence on a given problem. Sometimes everyone has all the collected evidence. After the outbreak of swine flu, epidemiologists in Great Britain used a common tracking database of cases, called QSurveillance [17]. Sometimes analysts have partial evidence. In the U.S., restrictions define which intelligence analysts can view which portions of intelligence data. One goal of this paper is to explore how the distribution of evidence influences the collaborative analysis.

When each analyst has all of the data or evidence, the demand for timely exchange of raw facts is minimal and discussion can focus on inferences and hypotheses drawn from the data. At the same time, having all the data raises the specter of information overload. To cope, analysts may discuss limited hypotheses to attain a common mental model. Although many writers argue that groups need a shared mental model [e.g., 4, 21], it can lead to confirmation bias. Thus, even in small groups with limited information to share, knowledge gains and improved performance from full information are seldom realized [29].

When analysts have only partial access to evidence, there is much more demand for information exchange; often the problem cannot be solved without it. For time-sensitive problems, valuable time will be spent simply making sure that everyone has the right information. To save time, analysts may decide to share lines of investigation or hypotheses, rather than raw data. For example, if a detective has noticed that many crimes take place near hospitals, he might share this observation with fellow detectives, rather than all his crime cases. If each analyst contributes a unique perspective, the analysts may debate alternative hypotheses, thereby avoiding confirmation bias. Still, key insights might emerge only when the collaborators link their ideas with the evidence. On balance, we propose that when analysts do not have all of the evidence themselves, they are likely to spend more time discussing hypotheses and relating them to the evidence than when they already have all the evidence.

Hypothesis 1: Pairs of analysts will more often solve the problem, discuss the problem more, and generate more hypotheses and better-supported hypotheses, when each partner has partial evidence than when each partner has all the evidence.

Information Visualization

Visualization techniques represent complex numerical and textual information in pictorial or graphical form and allow for visual exploration of data. Visualizations can help individuals spot anomalies and perceive patterns [22], increase the efficiency of information retrieval tasks and data analysis [35, 36], and promote feelings of community and foster discussion [38]. Researchers are pursuing visualization tools for collaboration such as CoVis [9], C-spray [32], CVD and Cave6D [23], TIDE [33], iScape [5], and COVISA [39]. However, little is known about how collaborations benefit from these systems, why they help, and whether there are limits to their benefits.

In their evaluation of CACHE [8], a system with visual data presentation for intelligence analysis, Billman et al. [3] report that distributed pairs using CACHE overcame *a priori* biases and did more effective data analysis. Mark et al. [25,26] reported that remote pairs with visualizations communicated more than collocated pairs did. Their results suggest that communication helps pairs take advantage of the visualization tool. From this work, we posit:

Hypothesis 2: Pairs of analysts with a visualization tool will more often solve the problem, will discuss the problem more, and will generate more hypotheses about the data and better-supported hypotheses, than analysts without a visualization tool.

Visualizations with All or Partial Evidence

If visualization tools provide the benefits we have discussed above, the degree of benefit may depend on the way in which evidence is distributed across members of a team. Although visualizations may be expected to improve hypothesis generation, discussion, and problem solving regardless of how evidence is distributed among analysts, these benefits may be reduced when the analysts each have all the evidence and reduced demand for information exchange and discussion.

Hypothesis 3: Visualizations will benefit collaborative analysis more when each partner has partial evidence than when each partner has all the evidence.

METHOD

We report analyses of data from an experiment designed as a two-level factorial, with two information conditions (half evidence vs. all evidence), and three visualization conditions (none, unshared visualizations, shared visualization). Participants worked in pairs randomly assigned to one of the three visualization conditions. We collected the data in the half evidence conditions for a previous study [1]. We subsequently collected data for the all evidence conditions to understand the significance of the distribution of information.

Participants

One hundred eighty total participants participated in the experiment, described as a "Detective Mystery Study" (84 female, 96 male; 55% U.S. born; age range 18-64, median age approximately 22). Eighty-eight percent of the participants were undergraduate or graduate students. Participants were paid $15 for their participation. They were told the experiment would last 1.5 hours. There were no demographic differences between the participants across conditions.

Procedure

Participants were seated apart, such that they could not see their partner or their partner's workstation. They role-played a pair of detectives of a police department, collaborating remotely to identify a possible serial killer. They had to work through many documents and reports to detect the serial killer. After working together on this task, they were each asked to complete two online reports on the results of their investigation, one on their serial killer analysis and another to report any other criminal activity they wanted to convey to the police department.

Participants in the visualization conditions were trained to use NetDraw (see Figure 1), the visualization tool adapted for this study. The no visualization control pairs were trained to use a spreadsheet that contained the same data. Participants were familiarized with the concepts of nodes and relationships, and they practiced on an example case using search and manipulating the diagram by location, time, and type of crime to give different perspectives on the evidence. Training took an average of 30 minutes.

After training, the pairs were left to work on the assignment for one hour. They were given an MSN Instant Messenger [IM] client and encouraged to use the client to talk with their partner. After an hour, or when the participants had completed their investigation and report, they each completed an online survey to elicit the evidence they used to identify the serial killer.

Serial Killer (SK) Task

The pairs' task was to identify a possible serial killer from the myriad of evidence the pair had on one current and six cold murder cases. The serial killer was responsible for four of the six homicides in a cold cases folder. Eight pieces of evidence, six within the cold case files and one in the open homicide case file, could be linked to the serial killer: similar blunt force trauma injuries to the victims; victims killed in the evening after they returned from work; victims rode the same bus route; victims lived near the same bus route; offender worked at a local hospital on the bus route; offender had been identified on the bus (alibi for a homicide witness); offender had been seen carrying a tool box on the bus. Identifying the serial killer required conceptually linking these disparate pieces of evidence from different cases rather than simply eliminating a defined group of suspects in one current case folder.

An open homicide case concerned the murder of a woman named Darlene Raffield. To solve this homicide, participants only had to examine the documents in one folder, review the alibis of witnesses, and evaluate their motives and opportunities to commit the crime. If a pair spent time on this case, they would be on the wrong track and have less time to complete the complex serial killer task. In pretesting, we found that individuals who spent more time on the Raffield homicide were less likely to identify the serial killer.

Distribution of Evidence

The evidentiary documents and reports were available online and could be opened, searched, put in different or new folders, and manipulated freely. To insure that sufficient screen space was available to examine multiple documents at once, the participants each had access to two 17" monitors placed side by side. Also, participants were given paper versions of the instructions and worksheets.

Participants had witness and suspect interview reports in the case files, coroner's reports, crime statistics by police district zone, a map of the zone and adjacent zones, a bus route map, and a police department organizational chart. Participants also could use a worksheet for recording dates, weapons, and other relevant evidence for each case, a suspect worksheet for recording different suspects, their connection to the victim, and alibis, and a timeline worksheet for recording when and where each crime took place, intended to support inter-case linkages.

In the *Half Evidence* condition, each member of the pair had half of the caseload and evidence for the serial killer on their computer. In the *All Evidence* condition, each member of the pair had all of the cases and documents.

Visualization Tool

Each pair was randomly assigned to one of three conditions, differing with respect to their use of a visualization tool. The tool enabled participants to see social and information network relationships in the data that linked names, places, events and objects, thereby providing a visual analysis perspective to identify the serial killer.

The tool was an adaptation of NetDraw v.2, a software application for drawing 2D social network diagrams, available online from Analytic Technologies. Social network diagrams are aptly suited for complex problem solving of the kind we used. The diagram connected over 50 unique names that represented how each person was connected to various other persons, and to view how they might be connected across cases.

Victims were represented in red and other persons, such as witnesses and suspects, in blue. Thick lines denoted strong ties, such as married people; thin lines denoted weak ties, such as two people who were observed at the same place at the same time.

Participants could freely manipulate and move the nodes within the screen, but they could not change the underlying relationships. Participants also could search or filter the diagrams based on a set of attributes to reveal people with common characteristics. Searchable attributes included police district zone, case affiliation, occupation, mode of transportation, time of crime, location of crime, weapon used to injure the victim, and the injured body part of the victim. For example, within the attribute weapon, the three options were handgun, blunt instrument, and poison. If handgun were selected, all victims who were injured by a handgun would be visible on the screen.

In the *No Visualization* condition, pairs did not have access to NetDraw. To ensure that they received the same information as others, they were given Microsoft Excel spreadsheets containing the same relationship information among the persons mentioned in the evidence documents. The names of people were arranged to form a matrix. Relationships in the matrix were represented by 2, 1, or 0, reflecting a strong tie, weak tie, or no relationship.

In the *Unshared Visualization* condition each member of the pair had access to NetDraw and an interactive and searchable social network diagram of their own evidence. They could not view their partner's visualization.

In the *Shared Visualization* condition, each member of the pair had access to NetDraw and an interactive and searchable social network diagram of all the evidence. This diagram could be manipulated and searched by both participants in the pair. The diagram was shared via a third computer using TightVNC, an open-source remote desktop software application. Effectively, this condition meant that, in the Half Evidence condition, each partner could see a diagram of all the evidence even though they only had direct access to half of the supporting evidence on their own computer. In the All Evidence condition, each partner not only had all the evidence on their computer, but also saw a diagram of all of the evidence.

Measures

We have three main sources of data, participants' final reports, their posttest surveys, and IM logs of their discussions.

Identifying the Serial Killer

Participants' correct identification of the serial killer was taken from their written reports. We were mainly interested in the success of the collaboration, so both members of the pair had to have named the serial killer for the pair to be coded as having collaborative successful performance. However, the results are essentially the same at the individual level.

Discussion Process

We calculated how much the pair communicated by counting the total number of IM words they exchanged during a session. We also coded participants' discussion topics, line by line, more than 8,700 lines of IM (see Table 1). An independent coder coded 7% of the data (Kappa = .71). All codes were at the individual level. Hypotheses were only counted the first time they were discussed, even if pairs revisited it after considering other hypotheses in between. The reason for this coding decision was that prior research suggests that the consideration of unique hypotheses, not the total number of times a hypothesis is mentioned, contributes to problem solving success.

Individual Characteristics

Prior research suggests that individuals' tendency toward cognitive reflection, as measured by a simple scale called the CRT, improves their ability to overcome confirmation bias (10). We used CRT scale scores as a control variable in our analyses. We also administered the NASA TLX scale, a measure of task workload [14].

RESULTS

We analyzed data from 90 pairs (180 participants), 15 pairs in each of the six conditions.

Identifying the Serial Killer

From our argument that distributed evidence leads partners to discuss and debate problems more deeply, we predicted that pairs whose partners each had only half of the evidence would perform better than those pairs in which both partners had all of the evidence. We also predicted that visualizations would help pairs solve the problem. Because the dependent variable, identifying the serial killer, is a discrete variable, the appropriate analysis is a logistic regression [16]. This regression assesses whether the independent variables predict the dichotomous outcome,

Topic	Definition	Example
Serial killer task	Pertains to solving the serial killer task or evidence pointing to the serial killer.	"I see a connection between 2 of my cold cases; they both involve a blunt object."
Clue Discussion	Discussion pertaining to one of the eight critical clues.	Detective A: "Hey, all of our victims ride the 500 bus." Detective B: "Ooh, good find!" or "That make sense, they all lived near the 500 as well!"
Hypothesis Discussion	Discussion in a new hypothesis is introduced or confirmed.	"I think these four blunt instrument victims are connected." "I feel like it is a suspicious man on the bus."

Table 1. Conversational coding scheme.

identifying the serial killer. We conducted analyses at the pair level. We found that performance depended on whether the pair had access to all of the evidence. In Figure 2, we see the results of the analysis, which support Hypothesis 3, the interaction effect.

In the Half Evidence condition, only 13% (*SE* = 12.5) of pairs in the No Visualization condition identified the serial killer, 46% (*SE* = 11.8) of pairs in the Unshared Visualization condition, and 60% (*SE* = 11.8) of pairs in the Shared Visualization condition identified the serial killer. Student's *t* tests show differences at the $p < .05$ level between No Visualization versus the Shared Visualization condition. In the All Evidence conditions, however, all three conditions performed comparatively poorly: 33% (*SE* = 9.01) of pairs in the No Visualization condition, 27% (*SE* = 13.3) of pairs in the Unshared Visualization condition, and 27% (*SE* = 13.3) of pairs in the Shared Visualization condition identified the serial killer (logistic regression Likelihood Ratio $\chi 2 = 9.3$, $p < .09$, $df = 5, 90$; Cramer's Phi = 0.35). The two visualization conditions in the Half Evidence condition significantly outperformed both All Evidence visualization conditions (logistic regression Likelihood Ratio $\chi 2 = 8.4$, $p < .05$, $df = 3, 120$).

Because so many pairs failed to identify the serial killer, we rated each participant's reports based on his or her progress towards solution on a four point scale: 0 for *unsolved*, 1 for *suspected pattern*, 2 for *suspected perpetrator,* and 3 for *correct solution*. We conducted an ANOVA with the solution as the dependent variable, evidence condition and visualization condition were between groups factors, and CRT scores were a control. (Non-integer degrees of freedom may occur in these analyses, see [24]). We found a significant effect by evidence condition on solution rates ($F[1, 82.36] = 4.57$, $p < .05$; Cohen's $d = 0.46$) and no effect by visualization condition. Individuals in the Half Evidence condition ($M = 1.87$, $SE = .12$) had significantly better solutions than those in the All Evidence condition ($M = 1.34$, $SE = .13$).

In summary, we found that visualizations did increase problem solving success as predicted, but only when evidence was distributed. We now turn to our analyses of the discussions to evaluate why shared and unshared evidence changed collaborative analysis.

Discussion Process

Total Talk

We predicted that pairs with half the evidence would discuss the problem more than pairs with all the evidence. We counted the total number of words each participant contributed to their IM discussion. We log transformed the data because they were skewed. In an ANOVA, the amount of total IM words was the dependent variable, evidence condition and visualization condition were between groups factors, and CRT scores were a control. As predicted, individuals in the Half Evidence conditions ($M = 446$, $SE = 20.9$) exchanged significantly more words with their partners than individuals in the All Evidence conditions ($M = 256$, $SE = 15.2$; $F[1, 80.8] = 28.9$, $p < .01$).

We also predicted that the visualization tools would increase discussion among pairs. Overall, there was no effect by visualization condition ($F[2, 80.8] = .06$, $p = .94$). However, the interaction effect between visualization and information condition showed a trend in the predicted direction ($F[2, 80.9] = 1.90$, $p = .16$). These results are seen in Figure 3.

How did the amount of discussion affect solutions? Overall, the number of IM words was significantly correlated with better solution rates ($r = .21$, $p < .01$). However, the importance of discussion varied by condition. In the two conditions where solutions were most likely, Half Evidence/Unshared Visualization and Half Evidence/Shared Visualization, words and solutions were positively correlated with higher solution rates ($r = .34$, $p = .06$; $r = .54$, $p < .01$, respectively), whereas in the other conditions the correlations were lower.

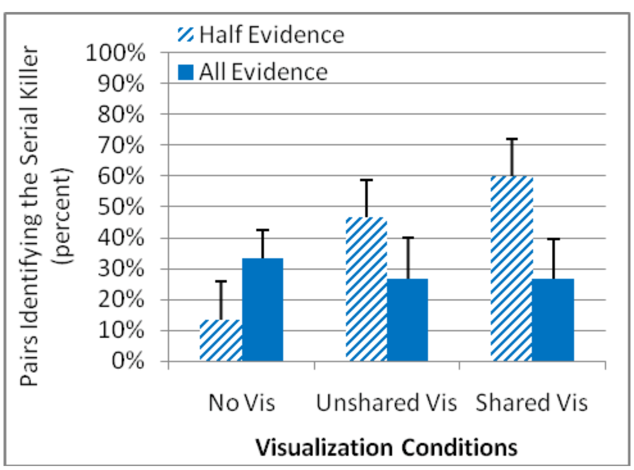

Figure 2. Percent of pairs solving the serial killer task by condition.

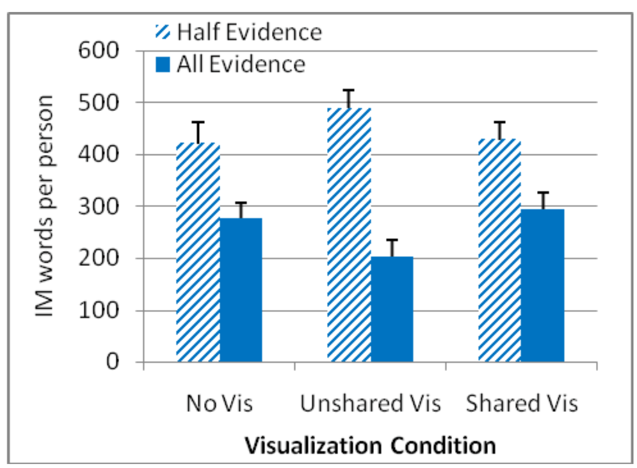

Figure 3. Average number of individual contributions of IM words by condition.

Discussion of Serial Killer

Total IM words and discussion of the serial killer, as measured by number of words (see Table 1) were highly correlated ($r = .76$, $p < .01$). We summed by individual the total number of words spent talking about the serial killer case and divided by the individual's total IM talk to control for individual variations of talk amount. In an ANOVA, the amount of serial killer discussion was the dependent variable, evidence condition and visualization condition were between groups factors, and CRT scores were a control. As predicted in Hypothesis 1, individuals in the Half Evidence conditions ($M = 256$, $SE = 15.2$) exchanged significantly more words with their partners about the serial killer case than individuals in the All Evidence conditions ($M = 153$, $SE = 15.0$; $F[1, 38.03] = 4.30$, $p < 0.05$). However, contrary to Hypothesis 2, those in the Unshared ($M = 198$, $SE = 18.4$) and Shared ($M = 203$, $SE = 21.5$) Visualization conditions did not talk more about the serial killer case than those without a visualization ($M = 218$, $SE = 19.1$; $F[2, 164] = .76$, *ns*).

Overall, discussion of the serial killer was significantly positively correlated with better solutions ($r = .47$, $p < .01$). This relationship was highest in the two conditions where there were the most solutions: the Half Evidence/Unshared Visualization and Half Evidence/Shared Visualization conditions ($r = .52$, $r = .57$, $p < .05$, respectively).

Discussion of Evidence

We reasoned that having a visualization tool and half the evidence would increase sharing of pieces of evidence critical to problem solving. We counted the number of critical pieces of evidence partners shared with each other in their IM discussion and compared that number to the number they recalled in the posttest survey. This analysis gave us a percent of evidence discussed of total critical evidence recalled for each individual. In an ANOVA, the percent of evidence discussed was the dependent variable, evidence condition and visualization condition were between groups factors, and CRT scores were a control. We only found a trend in the direction predicted in Hypothesis 1 (n.s.) for a main effect of type of evidence. Participants in the Half Evidence conditions discussed a higher percentage of the evidence with their partner ($M = 63.2\%$, $SE = 3.7\%$) than those in the All Evidence conditions ($M = 53.9\%$, $SE = 4\%$).

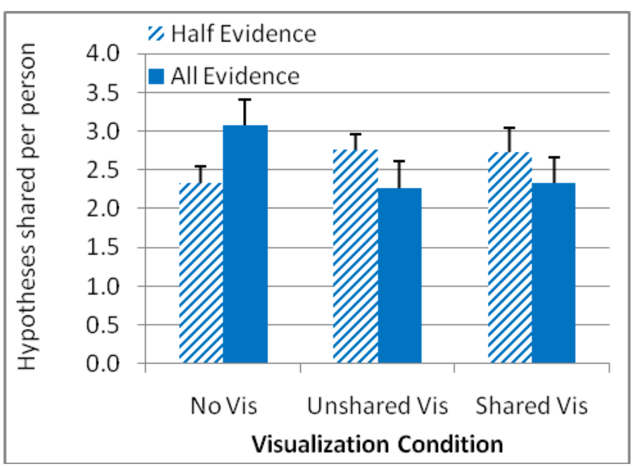

Figure 4. Average number of hypotheses shared per person across conditions.

Hypothesis Generation

We argued that generating more unique hypotheses would help pairs reach a solution. Consistent with this idea, the overall correlation between generating hypotheses and solutions was $r = .34$, $p < .01$. We hypothesized that having access to half the evidence would increase pairs' generation of unique hypotheses and that access to a visualization tool also increase would hypothesis generation. In an ANOVA in which the total number of unique hypotheses contributed to IM discussion by each individual was the dependent variable, evidence condition and visualization condition were between groups factors, and controlling for CRT scores, we found a marginal main effect by evidence condition ($F [1, 31.1] = 3.5$, $p = .07$) but no visualization main effect (see Figure 4). We also found a significant interaction effect between information and visualization condition ($F [1, 165] = 3.3$, $p < .05$), suggesting that visualizations helped in the Half Evidence conditions but not in the All Evidence conditions. In the Half Evidence conditions, Student's *t* tests show that those in the Unshared Visualization ($M = 2.77$, $SE = .2$) and Shared Visualization ($M = 2.73$, $SE = .3$) conditions discussed significantly more unique hypotheses than participants in the No Visualization condition ($M = 2.33$, $SE = .22$). For some reason, those in the All Evidence-No Visualization condition generated the most hypotheses. However, there was no correlation between generating hypotheses and solutions in this condition, suggesting that their discussion was comparatively fruitless.

DISCUSSION

We studied the impact of the distribution of information and a visualization tool on the process of collaborative problem solving. We found that using a visualization tool aids problem solving performance, but only when information is distributed between collaborative partners. This finding is contrary to the implicit assumption in much writing about data sharing that more access to data will aid collaborations. In this study, when both partners had access to all of the evidence, they could avoid the overhead associated with having to exchange information about the evidence. Despite this advantage, they performed more poorly and showed more evidence of confirmation bias than the half evidence conditions which performed better.

We speculate as to the reasons for these results. One explanation is that participants in the All Evidence conditions suffered from information overload. They had twice the number of text documents and far more evidence to look at. However, even in the Half Evidence conditions participants required more evidence than they themselves possessed to solve the serial killer problem, and so acquiring additional information from their partner (with the added overhead of communicating, representing, and storing that information) would presumably have increased their workload as much if not more than the All Evidence conditions. Furthermore, as measured by the NASA TLX scale on the posttest survey, participants in the All Evidence conditions did not report feeling a higher workload than participants in the Half Evidence condition.

The splitting of the evidence between partners also may have implicitly provided a social structure to the collaborative process. For instance, each of the partners can first go through his or her evidence and share his/her perspective with the partner. By contrast, partners with all the evidence may arbitrarily start sifting through the evidence and fail to consider the partner's viewpoint. Prior research has shown that assuming what the other partner knows leads to lower rates of collaborative success [31]. Significantly, participants in the All Evidence condition recalled as much evidence on the posttest as participants in the Half Evidence condition. However, they discussed a lower percentage of that evidence with their partner. This finding suggests that they lacked motivation to share information, perhaps because they assumed that their partner was aware of the same evidence.

Giving participants half the evidence may also have given them a sense of ownership and expertise about their own evidence. If people feel their contributions are important to collaborative success, they are less likely to show social loafing [20]. With both partners actively sharing ideas, there is a greater diversity of ideas within the pair, which can be associated with better collaborative outcomes [19]. Finally, a sense of differing expertise within groups helps reduce the tendency to focus on already shared information, which mitigates confirmation bias [37].

We will explore in more depth the explanation that participants in the All Evidence condition were taken off course by confirmation bias, as has been found by other research in collaborative analysis [7, 29]. One might think those with partial evidence might generate a narrow perspective, based on their own data, that would anchor their point of view. However these pairs knew at the same time that the partner had relevant evidence. This knowledge could be crucial and a big reason why those in the Half Evidence/Shared Visualization condition did so well. In that condition, partners had only half of the evidence but they could see a diagram of all the evidence, including their partner's on the screen. This knowledge would have elicited conversation about the problem and a search for evidence. This conversation may have forced confrontation with disconfirming evidence for incorrect hypotheses.

In order to pursue the confirmation bias explanation further, we performed a detailed tracing of the discussions in the Shared Visualization conditions, where the differences between the Half Evidence and All Evidence conditions were most stark. Figure 5 shows a mapping of all discussions in those conditions. Each dot represents one pair in either the All Evidence (whole circles) or Half Evidence (half circles) condition. Orange (or light grey in black and white) dots represent failures to solve the case and blue dots (dark grey in black and white) represent identifying the serial killer. In the far right path, all pairs in the All Evidence conditions who started with the irrelevant Raffield case remained stuck there, whereas 5 out of the 7 pairs in the Half Evidence conditions successfully moved on to solve the serial killer case. In the middle path, pairs in the Half Evidence condition who saw a serial killer pattern first in their process then understood a crucial connection between cases whereas a majority of those in the All Evidence conditions who saw a pattern did not successfully identify a connection. In the far left path, more All Evidence pairs noticed a connection between two cases than did those in the Half Evidence conditions, but seeing this connection only translated to successfully understanding what the connection meant and identifying the serial killer 50% of the time.

This analysis suggests why we believe confirmation bias plagued those in the All Evidence conditions, and visualizations did not help them. If they started off on the wrong path, they were more likely to stay there, and even if they noticed an interesting clue, such as the connection between two cases or the serial killer pattern, they did not debate the data sufficiently to come to a correct solution.

Limitations

Although this study contributes to understanding how the distribution of information and visualizations can affect collaborative analysis, we have studied only one analytic task, limiting generalizability. Also, we studied people who had not worked together. Collaborators may build

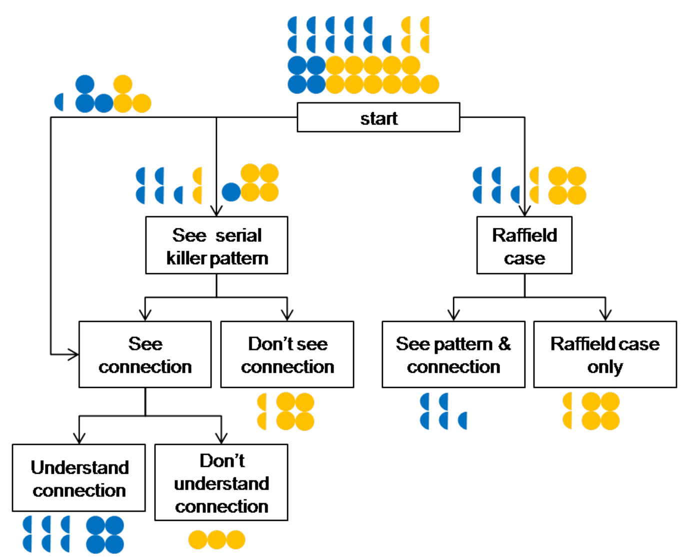

Figure 5. Process map of integrated visualization pairs. Full circles represent All Evidence condition pairs and half circles, Half Evidence condition pairs. The color blue designates pairs who solved the serial killer case while orange designates pairs who did not solve the case.

experience working with one another, improving their communication (however, this experience seems not to reduce confirmation bias [15]). Also, our participants used IM, whereas real-world collaborations most likely rely on more than just one form of communication, including audio and video channels. Finally, in real world environments, analysts are highly trained in the use of knowledge and visualization tools, whereas our participants may have suffered from inadequate experience despite some training in its use.

Design Implications

The results of this study may have implications for the design of collaborative visualization tools. Pairs did not do very well when each partner had access to all the evidence, even with an interactive diagram linking the data in logical ways. We propose that tools for sharing intermediate steps in analysis could aid collaborative investigative analysis process by helping groups overcome information overload, confirmation bias, and social psychological barriers to effective analysis such as social loafing.

Surfacing traces. One possible design direction is to address the barriers to analysts deciding how to partition the task. Only three pairs in the All Evidence condition proposed splitting up the evidence and examining parts of it in more depth, and in all three cases they quickly abandoned the idea. One way to facilitate analysts looking at different, relevant parts of the evidence might be to surface traces of their paths through the data. That is, as analysts explore the data, they create traces that can be aggregated and displayed, such as how long they examined a suspect or whether they discussed the suspect with others. This idea builds on Gutwin's work [13], applied to analytic tasks.

One possible design illustrating this approach is shown in Figure 6, which describes three detectives, Alice, Hao-Chin, and Carlin, working on the serial killer task. Looking at the trace-surfacing, Carlin sees that Hao-Chin has spent time looking at Wayne Millican, and that he has annotated that Millican was carrying a toolbox on the bus when interviewed for an unconnected case. This awareness may drive Carlin to spend his time on other pieces of information that have not been as well investigated, and he is likely to have an eye out for the same suspect. While such designs can be unsuitable for specific quantitative calculations [6], they may support the general awareness of collaborators' activities.

Sharing categories. In our study, pairs that shared more intermediate structures in the forms of hypotheses or clues had a higher likelihood of solving the problem. However, it is difficult, especially with limited communication channels such as IM or even voice, to easily share these mental structures. Visualizations might help.

Analysts may use many different ways of categorizing evidence as they explore it, from simple categories such as "victims killed with blunt instruments," to more ad hoc categories such as "people who ride the bus to work but not back." By choosing what information should be included in a category, collaborators focus and highlight different aspects of the information. Being able to visualize the aggregated categorization structures of many individuals could help collaborators better understand the mental representations of their collaborators, make sense of the way others are grouping data, and induce higher-order schemas (such as the presence of a serial killer).

Studies of how people represent and use concepts highlight that categories are often flexible, evolving, ad-hoc, and

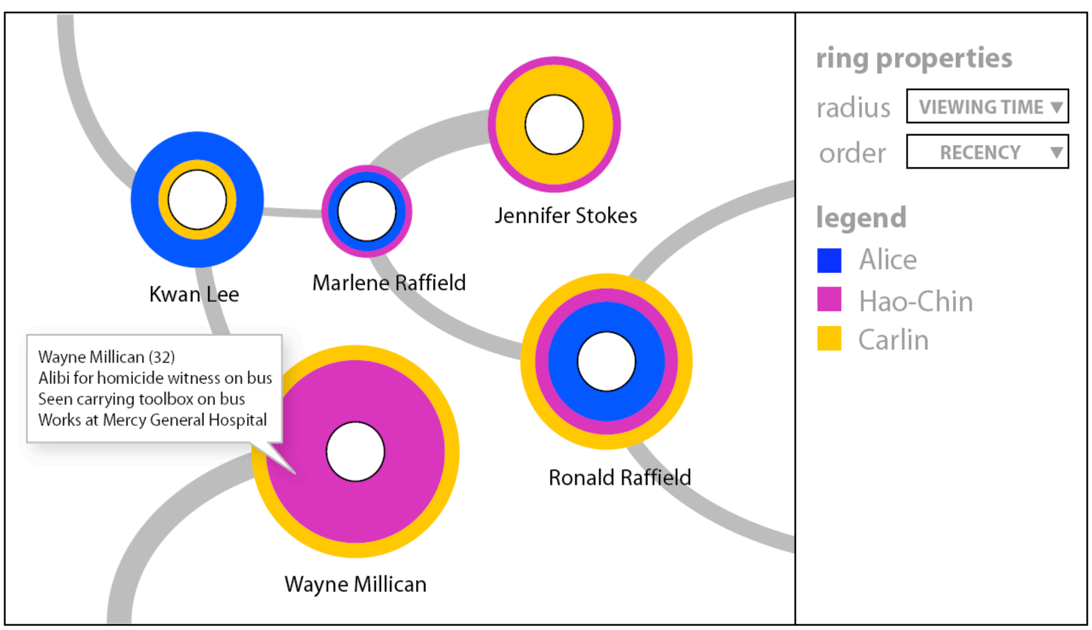

Figure 6. Surfacing analysts' traces. Investigating Wayne Millican, Carlin sees that Hao-Chin has already spent significant time examining and annotating him.

theory-driven rather than determined by static features of data [2,30,39]. Thus, there is no top-down correct way to categorize data; analysts will need to organize and reorganize data in many different ways. We thus suggest that visualizations should effectively enable analysts to flexibly organize information and share those organizations with each other.

Sharing hypotheses. Once analysts have explored categorization structures they may build up more sophisticated hypotheses about what is going on. A hypothesis or schema can represent a set of relations between items, such as that there is a serial killer in the region; animals from a set of farms may be the source of new outbreak, or that a specific person is involved in a conspiracy. Inducing such hypotheses is difficult, as they require the integration of many, often disparate pieces of information [11,18]. Groups can promote fruitful problem solving when individuals generate hypotheses that others can then build on. For example, two detectives might combine their hypotheses in independent cases to identify a serial killer. Supporting coordination is especially important for hypothesis sharing, as collaborators need to share representations and mental models of the information space, suggesting that visualizing annotations and hypotheses of others could be highly beneficial. For instance, in the detective example above, one detective might note an anomaly in his case, which another detective could then use to induce a higher level schema across cases (i.e., that there is a serial killer).

An important research question to answer with respect to all these design ideas is whether sharing intermediate structures increases the danger of confirmation bias. Prior research has shown that a diversity of perspectives in a collaboration can help prevent confirmation bias [8,19,28]. Thus visualizations should allow inconsistencies in the evidence to be viewed, so that apparently similar hypotheses can be compared.

CONCLUSION

Visualizations improved remote collaborators' performance over the control condition but this improvement depended on information load. When each partner had all the evidence, discussion flagged, pairs discussed fewer hypotheses, and they persisted on the wrong hypothesis—i.e., confirmation bias. These results imply that the design of remote collaboration systems should incorporate awareness of intermediate processes important to collaborative success.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation Grant, IIS-0325047. We thank Peter Scupelli for help developing experiment materials.

REFERENCES

1. Balakrishnan, A.D., Fussell, S., Kiesler, S. (2008). Do visualizations improve synchronous remote collaboration? *Proc. CHI 2008.* NY: ACM.
2. Barsalou, L. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211-227.
3. Billman, D., Convertino, G., Pirolli, P., Massar, J. P. and Shrager, J. (2005). Collaborative intelligence analysis with CACHE: Bias reduction and information coverage. Unpublished manuscript. Palo Alto Research Center, CA. Retrieved from http://cscl.ist.psu.edu/public/users/ gconvert/mypapers/hcic2006_BillmanEtAl.pdf.
4. Blockeel, H., & Moyle, S. (2002). Collaborative data mining needs centralised model evaluation. In Lavrac, N., Motoda, H. and Fawcett, T. (Eds.) *Proc. ICML-2002 Workshop on Data Mining Lessons Learned* (pp. 21-28). Leuven, Belgium.
5. Börner, K. (2001). iScape: A collaborative memory palace for digital library search results. *Proc. CHI 2001* (pp. 1160–4). NY: ACM press.
6. Cleveland, W.S., & Schmieg, G.M. (1987). The elements of graphing data. *American Journal of Physics*, *55*(8), 767.
7. Cook, M. B., & Smallman, H. S. (2007). Visual evidence landscapes: Reducing bias in collaborative intelligence analysis. *Proc*. *Human Factors and Ergonomics* (Vol. 51, pp. 303-307).
8. Convertino, G., Billman, D., Pirolli, P., Massar, J. P., & Shrager, J. (2008). The CACHE study: Group effects in computer-supported collaborative analysis. *Proc. CSCW* (pp. 353-393). NY: ACM Press.
9. Edelson, D., Pea, R., Gomez, L. (1996). Constructive in the collaboratory. In B.G. Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design*. Englewood Cliffs, NJ: Educational Technology Publications.
10. Frederick, S. (2005). Cognitive reflection and decision making. *J.of Econ Persp*, 19(4), 25-42.
11. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155-170.
12. Greif, I. & Sarin, S. (1987). Data sharing in group work. *ACM Transactions on InfoSys*, 5, 187-211.
13. Gutwin, C. (2002). Traces: Visualizing the immediate past to support group interaction. In *Graphics Interface* (pp. 43-50).
14. Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale. In P. A. Hancock & N. Meshkati (Ed.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier.

15. Heuer, Jr., R. J. (1999). *The psychology of intelligence*. Washington DC: Center for the Study of Intelligence, Government Printing Office.
16. Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. NY: John Wiley & Sons.
17. www.qresearch.org/Public/QSurveillance.aspx
18. Hummel, J., & Holyoak, K. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psych Review*, *110*(2), 220-264.
19. Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in work groups. *Adm Sci Quarterly*, 44, 741–763.
20. Karau, S. J., & Williams, K. (1993). Social loafing: A meta-analytic review and theoretical integration. *J. Pers & Soc Psych, 65*(4), 681-706.
21. Kozlowski, S. W. J., Ilgen, D. R., & Klimoski, R. (2006). *Enhancing the effectiveness of work groups and teams*. Blackwell Publishers.
22. Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65-100.
23. Lascara, C., Wheless, G., Cox, D., Patterson, R., Levy, S., Johnson, A. E., & Leigh, J. (1999). TeleImmersive virtual environments for collaborative knowledge discovery. *Advanced Simulation Technologies*. San Diego, CA.
24. Littell, R., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.
25. Mark, G., Carpenter, K., & Kobsa, A. (2003a). Are there benefits in seeing double? A study of collaborative information visualization. *Proc. CHI 2004* (pp. 840-841), NY: ACM Press.
26. Mark, G., Carpenter, K., & Kobsa, A. (2003b). A model of synchronous collaborative information visualization. *Proc. Info Vis 2003* (pp. 373-381). NJ: IEEE Press.
27. Mark, G., Kobsa, A., & Gonzalez, V. (2002). Do four eyes see better than two? Collaborative versus individual discovery in data visualization systems. *Proc. InfoVis 2002* (pp. 249-255), NJ: IEEE Press.
28. Mohammed, S., & Dumville, B. (2001). Team mental models: Expanding theory and measurement through cross-disciplinary boundaries. *J. Org Behavior*, 22, 89-106.
29. Mojzisch, A., & Schulz-Hardt, S. (2005). Information sampling in group decision making: Sampling biases and their consequences. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 299–326). Cambridge: Cambridge Univ. Press.
30. Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289-316.
31. Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175-220.
32. Pang, A., & Wittenbrink, C. (1997). Collaborative 3D visualization with CSpray. *Computer Graphics and Applications, IEEE*, *17*(2), 32-41.
33. Sawant, N., Scharver, C., Leigh, J., et al. (2000). The tele-immersive data explorer: A distributed architecture for collaborative interactive visualization of large data-sets. *Proc. Immersive Projection Technology Workshop,* Ames, Iowa.
34. Shepperd, J. A. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin*, *113*, 67-67.
35. Shneiderman, Ben. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proc. Visual Languages 1996* (pp. 336–343). NJ: IEEE Press.
36. Stasko, J., Catrambone, R., Guzdial, M., & McDonald, K. (2000). An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human Computer Studies*, *53*, 663-694.
37. Stasser, G., Vaughan, S. I., & Stewart, D. D. (2000). Pooling unshared information. *Org Beh & Human Decision Processes*, *82*, 102-116.
38. Viégas, F., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. *Proc. CHI 2004* (pp. 575-582). NY: ACM Press.
39. Wittgenstein, L. (1953). Philosophical investigations. Philosophische Untersuchungen.
40. Wood, J., Wright, H., & Brodlie, K. (1997). Collaborative visualization. *Proc. Vis 1997*. NY: IEEE Press.