HCC: Medium: Collaborative Research: Scaling Collaboration over Data and People

Project Description

As the recent meltdown of the financial markets has made clear, the world has become significantly interconnected. Individuals and organizations not only have access to enormous amounts of information but their tasks and lives are increasingly affected by what others do with this information. For the good, sites such as Wikipedia demonstrate how hundreds of thousands of contributors can pool their resources to create an encyclopedia of knowledge (Kittur & Kraut, 2008). Collaborations in teams and alliances now try to solve problems as disparate as investigating related homicides, documenting financial corruption, predicting elections, determining the sources of global warming, making field-spanning scientific discoveries, and tracing outbreaks of disease. Many endeavors have become more collaborative than not. In scientific research, for instance, teams increasingly dominate productivity in fields ranging from chemistry to psychology (Wuchty, 2007). Intelligence analysis (Heuer, 1999) and business innovation (Baron, 2006) also depend on whether individuals, groups, and organizations can integrate their knowledge, expertise, and insights.

Here, we propose basic research in collaborative investigative analysis with intended application to domains such as financial and political investigation, criminal and intelligence analysis, and epidemiology. By collaboration, we mean more than information or data sharing (Greif & Sarin, 1987). We also mean shared sensemaking—whether people working together can filter, disambiguate, and connect data, build schemas, and integrate information to identify and solve problems, make decisions, and take action (Nickerson, 1992). Collaborators' goals might be to trace money laundering, to identify potential perpetrators of computer network attacks, or to identify the reason for a town's high cancer rates. To achieve each of these goals, collaborators sift through evidence and explore hypotheses, trying to create logical linkages that point to an appropriate problem definition, trail of evidence, or insight. Great strides have been made in understanding, training, and supporting decision making in comparatively small groups (Kozlowski & Ilgen, 2006). However, much less progress has been made in understanding and improving investigative analysis in groups, particularly when the group encounters large amounts of data and there are large numbers of collaborators. Collaborative analysis across large amounts of data and many people is problematic, and the underlying cognitive and social processes are not well understood.

Consider the following criminal case. A police department in North Carolina had considerable information about the circumstances surrounding local murders but did not realize in this myriad of data that a serial killer was on the loose. Henry Wallace had already strangled at least five young women, and he would kill four more women, when he was finally arrested for shoplifting at a mall and released (State of North Carolina v. Henry Louis Wallace, 2000). Wallace was eventually identified as a serial killer when a detective noticed strangulation in deaths only two weeks apart and told colleagues, who looked for and then found Wallace's palm print on the car of a victim. As this case illustrates, significant breakthroughs in detective work often come about when detectives notice and discuss disparate and sometimes unlikely or competing hypotheses; colloquially, they "connect the dots." Similarly, in intelligence analysis (Heuer, 1999), business innovation (Baron, 2006) and scientific research (Klahr & Simon, 1999; Simonton, 2003), success may hinge on collaborators associating information that others have not noticed or thought unimportant.

Preliminary work

We have conducted initial research on collaborative analysis by comparing the value of collaborative vs. individual analysis in an experimental paradigm. We use a criminal investigative analysis task called the *Serial Killer Task.* In this task, based on actual cases (e.g., Kraemer et al., 2004), participants role-play detectives in the homicide unit of a local police department. Evidence pointing to a serial killer is scattered, for instance, on linkages among victims who ride a particular bus route. Participants read and evaluate assorted documents, including witness and suspect interviews, coroner's reports, crime statistics, a map of the zone and adjacent zones, a bus route map, and a police department organizational chart. In our first experiments (Balakrishnan et al., 2008, in press; Kiesler et al., 2008), participants worked either individually or synchronously with a partner through Instant Messaging [IM]. Across a series of studies, we manipulated how information was distributed across participant pairs, whether participants could use structured ways of organizing and reporting intermediate reasoning processes (e.g., an MO worksheet, timeline), and simple data visualization tools in the form of social network diagrams. We discovered that visualization tools improved analysis overall, but collaborative analysis was surprisingly *less* successful than individual analysis. In dyads, twice the intellectual power can be applied to the same data as when individuals do the analysis, but dyads did not perform better than solo analysts looking at the same amount of data and with comparable tools, and in many cases, they performed worse. Dyads also took longer to reach a conclusion. Our studies suggest that scaling up to just two people incurs coordination costs or "process losses" (Shepperd, 1993). Simply sharing information does not seem to help analysts reach the critical insight they need. Our finding fits with other work. For instance, Blockeel and Moyle (2002) report that barriers to analysts synchronizing their mental representations reduce the potentially added value of collaboration.

In a parallel effort, we have built alternative collaborative visualization systems. *Sense.us* (Heer et al., 2007) is a web site supporting asynchronous collaboration across a variety of visualization types via view sharing, linked discussions, and graphical annotation. User studies revealed emergent patterns of social data analysis, cycles of observation and hypothesis, and the complementary roles of social navigation and data-driven exploration. These results suggest benefits for asynchronous forms of collaborative analysis in which individuals pass their contributions to others (c.f., Benbunan-Fich et al., 2003). Further enhancements include scented widgets (Willett et al., 2007)—user interface controls that visualize traces of social activity to enhance collective information foraging—and data-aware annotation techniques (Heer et al., 2008) that support retargeting across representations of the underlying data. In controlled studies we found that these techniques can help analysts allocate their attention more effectively (e.g., by surfacing neglected data regions) and share annotations across diverse visual representations. While such tools provide a substrate for shared analysis, more research is needed to understand how to better facilitate and scale the integration of analytic insights (Heer & Agrawala, 2008). Our understanding of the design space of social data analysis tools is limited and, as a result, tools for social sensemaking remain in their infancy.

One way to improve collaborative investigative analysis, which we explore in the proposed work, is to automate or support analysts' sharing of intermediate products of their analysis work, such as their categorizations of data, their hypotheses, or tacit traces of their analytic behavior, such as the information they looked at and their paths through the data, given off as the analysts go about their work. Our major challenge is not only to create such tools, and to test their usefulness in collaboration, but also to assess whether they scale to large amounts of data and large collaborations.

Intellectual merit

The proposed research will improve our understanding of the complex cognitive and social coordination activities required in collaborative investigative analysis, and lead to new tools supporting this coordination. Our theoretical framework based on cognitive limitations and biases and social processes, exacerbated by scaling data and people, will guide new research on analysis and the potential of visualization tools. The research involves undergraduate, graduate, and professional students, and will result in their further training and education in interdisciplinary research. We propose interesting new educational activities, and have had experience in doing so in the past.

Broader impacts

This project has the potential to improve collaborative investigative analysis in many fields of critical importance to society, including criminal justice, intelligence, science, and epidemiology. Our results will provide new visualization tools for analysts in these areas, recommendations for organizational practices to improve the quality of collaborative analysis, new methods for training professional analysts to solve complex, interconnected problems, and new learning tools for graduate programs in fields such as epidemiological analysis and criminal justice.

Research team

The team consists of experts in organizational behavior (Kiesler), collaborative analysis (Kiesler, Fussell, Heer), computer-mediated communication (Fussell), information visualization (Heer), and information aggregation (Kittur). The PIs have a track record of successful interdisciplinary collaborations and have collaborated together on prior NSF-funded projects. Kiesler and Kittur are co-authoring chapters for a book on theory in online community design. Fussell and Kiesler have co-advised three doctoral students.

Scaling Collaboration

In collaborative investigative analysis, we face two issues of scale, data and people. Both of these issues complicate theory and the development of tools.

Scaling data

In hundreds of small group experiments, including our own, participants have worked with comparatively little data, for example, 40 informative items or simple documents per person in a three-person group (e.g., Greitmeyer et al., 2006). By contrast, real world tasks often involve what is commonly understood as "information overload." Consider for example, the thousands of data pieces that a homicide detective gathers and can consider in a portfolio of cases, and in addition, the various statistics and analyses from computers that the detective has at hand. If there are 10 detectives and they share their expertise and information with others, whether directly or through a database, the sheer amount of information will be overwhelming to them (see, for example, Boh, 2007). Even if the detectives can identify the unshared information each has, the group as a whole will have more information than it can analyze. In addition, the individual analyst will now have more information than he or she started with, leading to even greater overload. Thus, in complex environments with large amounts of data, information exchange alone is unlikely to enhance analysis.

One strategy a group of analysts can use to cope with immense amounts of data is to agree to focus on certain hypotheses or categories of data (e.g., male suspects, murder weapons) in one way (e.g., pairings of men of a given age with use of the same weapon). That is, they may use a common framework or shared mental model to approach the data. Although many writers argue that groups need a shared mental model to succeed (e.g., Kozlowski & Ilgen, 2006), too much overlap can lead to cognitive tunneling (Cook & Smallman, 2007; Woods & Cook, 1999) whereby the collaboration fails to discuss unshared information (Stasser & Titus, 1987) and follows an incorrect path (e.g., looking closely at all young male suspects). Cognitive tunneling leads to confirmation bias (Nickerson, 1998), that is, failing to consider alternative hypotheses (e.g., looking at older suspects).

Another strategy the group can use is for each detective to begin with a more prolonged period of individual analysis, sharing his or her lines of investigation or hypotheses only after some preliminary conclusions have been drawn (Dugosh et al., 2000). For example, if one detective has noticed that many crimes take place near hospitals, he might share this general observation with fellow detectives, rather than sharing all the available information about each individual crime. This strategy has the potential to reduce information overload and to increase creativity because each detective is contributing unique hypotheses or other intermediate products of the analysis process rather than raw data. This deliberate information pooling strategy, however, can be problematic in two ways. First, some key insights may not be evident within individual detectives' portions of the data; they emerge only when the data are combined with others' data. For example, if each detective is working on one crime that took place near a hospital, the fact that this is a general pattern across all detectives' data may not be obvious. Second, cognitive biases, social pressures, and organizational disincentives can make analysts unwilling to share tentative hypotheses. For example, analysts might worry they would be punished for false positives and wasted resources. For these reasons, even in small groups with limited information to share, improved performance from deliberate information pooling are seldom realized (for an overview of this literature, see Mojzisch & Schulz-Hardt, 2005).

Scaling people

Fred Brooks famously said, "Adding manpower to a late software project makes it later" (Brooks 1975). In investigative analysis, as well, a major issue of scale has to do with what happens when we increase the number of people involved in a collaboration. Although increasing the number of contributors to analysis potentially increases the likelihood that someone has a key insight or expertise needed, the evidence suggests that adding more people actually reduces efficiency and effectiveness. More people increases the costs of coordination—effort to meet with others, time spent learning who knows what, contextualizing information, deciding who should meet with whom, agreeing on the problem or tasks, keeping up with what is going on, and coming to consensus. Coordination costs in turn reduce the efficiency and effectiveness of a group (Steiner, 1972; Schulz-Hardt et al., 2006). Adding more people also can yield diminishing returns because some members of the collaboration feel distant from others, slack off, or begin to feel unmotivated or unappreciated; losses are especially evident when the task has strong interdependencies (Hill, 1982; Karau & Williams, 1993; Shepperd, 1993).

The people in a collaboration may belong to two or more different organizations. Analyst teams comprised of people from different organizations are becoming more common, given the complexity of problems analysts face in solving global crime, pandemics, and other geographically and organizationally distributed problems. Collaborations with members from multiple organizations, however, often suffer from increased coordination costs and social conflict due to members' divergent organizational procedures, routines, cultures, and incentive structures (e.g., Cummings & Kiesler, 2005, 2007; Hinds & Mortensen, 2005; Katz & Te'eni, 2007). When such collaborations are also carried out over distance, the probability of serendipitous encounters drop drastically (Boh, 2007). It also becomes significantly more difficult to locate individuals with a specific piece of knowledge. Hansen and Nohria (2004, p. 24) refer to this search as the ''needle-in-a-haystack problem," noting that ''it is nearly impossible to connect the person who has the expertise with the person who needs it.''

Need for collaborative visualization tools that scale

Visualization techniques represent complex numerical and textual information in pictorial or graphical form and facilitate exploration of data (Andrews & Heidegger, 1998; Shneiderman, 1996; Wattenberg, 1999). By removing the burden of mentally consolidating disparate information, such holistic representations of large amounts of data can help individuals spot anomalies, perceive patterns, and thus improve their problem solving success (Larkin & Simon, 1987). Visualization tools reduce task completion time and increase productivity on many information retrieval tasks (Hendrix et al., 2000; Stasko et al., 2000; Veerasamy & Belkin, 1996) and are commonly used to support intelligence analysis (Stasko et al., 2008; i2, 2009; Palantir, 2009; Wright et al., 2006; CoMotion, 2009). Information visualizations also can promote feelings of community and foster discussion in "wiki" websites (Viégas, Wattenberg, & Dave, 2004).

The importance of collaborative investigative analysis in many real world tasks has spurred significant research of visualization tools that support collaboration. Early systems such as CoVis (Edelson et al., 1996), C-spray (Pang & Wittenbrink, 1997), CVD and Cave6D (Lascara et al., 1999), TIDE (Sawant et al., 2000), iScape (Börner, 2001), COVISA (Wood et al., 1997), and DIVA (Brewer et al., 2000) focus primarily on system control issues, e.g., enabling users to manipulate a shared visual interface in real-time. More recent research attempts to support not just multi-user interaction, but collaborative sensemaking of visualized data, typically through textual discussions linked to visualization states (Heer et al., 2007; Viégas et al., 2007; Eccles et al., 2007) and aggregation of individually-constructed artifacts such as network diagrams or measures of evidence quality (Billman et al., 2005; Convertino et al., 2008, Brennan et al., 2006; Keel, 2007). While promising, such systems do not yet scale to large data sets or numbers of collaborators. We believe that such scalability requires systems that not only handle large databases, but also represent the process and intermediate products of sensemaking—such as categories and hypotheses—in a format amenable to both manual and computational analysis.

Much remains to be understood about the conditions under which groups optimally benefit from collaborative visualization, particularly in regards to higher-level sensemaking tasks. Some research suggests there are improvements in performance on certain analytical tasks, especially when the visualization is easily understood, the tasks require little coordination among collaborators, and the data and number of collaborators are constrained (Mark, Carpenter, & Kobsa, 2003a, 2003b). Visualization systems that minimize the overhead of planning and coordination among people lead to better group performance than systems with high coordination costs (Mark, Kobsa, & Gonzalez, 2002).

Only a few studies systematically examine the effectiveness of visualizations on such highly complex tasks as investigative analysis (e.g., Balakrishnan et al., 2008, in press; Convertino et al., 2008). In our serial killer studies mentioned above, we ran conditions in which participants were provided with social network diagrams to help them analyze the data. These visualizations improved the performance of both individuals and dyads; for example, a visualization integrating two sets of data held by different analysts improved their performance from a 25% solution rate (per pair) to a 67% solution rate. However, dyads did not overcome the collaborative challenge in sharing data, namely, that even with a diagram linking the data in logical ways, there is still too much to consider and there are coordination costs in discussing the data (Balakrishnan et al., 2008; Kiesler et al., 2008). Furthermore, when every analyst had all the data, performance in dyads seemed particularly prone to cognitive tunneling (Balakrishnan et al., in press). For a group to benefit from collaborative visualizations, tools must aid coordination cognitively and socially rather than be dependent on the pre-existing coordination of the group.

In summary, we call for research that will lead to a fundamental understanding and better support for collaborative investigative analysis in critical domains such as criminal and intelligence analysis. There is considerable research towards support for small groups, such as helping groups set goals, keep on task, and maintain awareness of colleagues (e.g., Pritchard, 1995; Sawyer et al., 1999). In addition, research in information aggregation has made strides toward helping individuals and groups retrieve and share large amounts of information. This work has been especially useful when the task can be modularized (as in Wikipedia) and individuals can make individual contributions. We propose to build on this prior work to understand and improve collaborative investigative analysis, which demands that information not just be shared, but also integrated. We propose developing tools, research, and theory that extend prior work in a qualitatively different direction. We hypothesize that tools for sharing the intermediate products of analysis will transform the collaborative investigative analysis process by helping groups overcome information overload, cognitive tunneling, and social barriers to investigative analysis. We also propose that such tools may scale better than sharing raw data.

Research Goals

Solving the problems faced by analysts collaborating on large datasets requires us to expand models of investigative analysis beyond data and relations among data to the collaborators and relations among them. It requires us to scale up the study of collaboration as it has been studied in social psychology to larger amounts of data and larger groups of people, including dispersed groups. It requires us to create tools that help groups of people coordinate their analysis work at different stages of analysis. We believe this approach will reduce coordination costs that pose barriers to information sharing and will help collaborators use amounts of information that greatly exceed individual cognitive capacities. Specifically, we have three key research goals:

Understanding how the process and results of collaborative analysis change as we increase the scale of data and people involved. Most theories of collaborative investigative analysis have been based on studies of small groups coping with manageable amounts of data. These theories will need to be extended and modified to fit the case of many analysts, sometimes from different organizations, collaborating on vast amounts of data. We need to understand how to best design tools that scale to larger groups of analysts and large amounts of data.

> *Exploring optimal cognitive and social conditions for collaborative investigative analysis*. Collaborations are social organisms that, over time, develop a particular culture and cognitive structure. We will address questions that remain unsolved in the literature: What degree of overlap in mental models leads to the best analytic success, with the least cognitive tunneling? What level of information should be shared: raw data, categories, or hypotheses? Should sharing be explicit, as through tags or speech; implicit, as through traces of analytic behavior; or both? How can computation facilitate matching and integration? How does collaboration across social boundaries affect trust and the willingness to share?

> *Designing and testing visualization tools for collaborative investigative analysis*. Visualizations are efficient ways of organizing information, a key component of investigative analysis. How can analysts benefit from the analytic work of others? We will examine how visualizations can help analysts make use of each others' analysis behavior, and whether these tools scale to large datasets and large collaborations.

We approach these three research goals in two highly intertwined research activities: development of visualization tools for collaborative investigative analysis, and behavioral studies of analysts using these tools.

Research Plan

Activity 1: Developing tools for collaborative investigative analysis
Our approach to developing tools for collaborative visualization focuses on ways in which analysts can benefit from seeing the activity and intermediate products of others. To illustrate our approach we present a hypothetical scenario based on the *Serial Killer Task* described above. This simplified scenario nonetheless illustrates ways how collaborative visualizations could help analysts share their categorizations of data and their hypotheses, and follow the analytic behavior of others. We then raise some challenging research questions that arise from these ideas.

Scenario

Suppose three analysts collaborate to identify suspects for further questioning. Using a visualization of category structures, Carlin sees that 4 of the 7 victims were killed with a blunt instrument. Intrigued, she uses this categorization structure as a source to look for related structures, finding one Alice had previously created for "bus route 22x riders". Looking at the overlap of the two, she finds that all of the four blunt instrument victims were also bus riders (see Figure 1). Thinking of a possible connection, she annotates the view that there might be a serial killer involved and adds it to the shared pool.

Alice, notified that someone has used one of her categorization structures, finds Carlin's combined view. She decides to try to find suspects who might be related. She comes up with a new hypothesis, and, using the underlying graph clustering algorithm, generates people who are linked to the victims (see Figure 2a). From this view, she finds two likely suspects: Wayne Millican and Ronald Raffield. She saves this view as a hypothesis, annotating that both are worthy of further inspection.

Looking at the trace-surfacing view (see Figure 3), Carlin sees that Hao-Chin has spent some time already looking at Wayne Millican, and that he has annotated that Millican was carrying a toolbox on

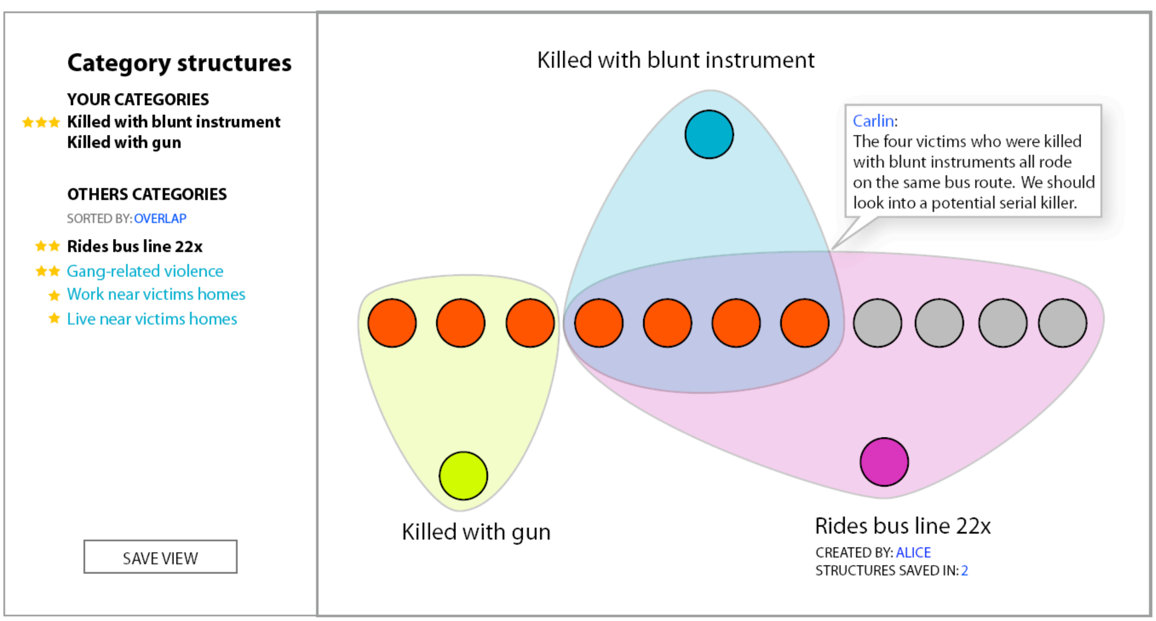

Figure 1. Sharing categorization structures. By combining her structure with Alice's, Carlin notices that all victims (in red) killed by a blunt instrument also rode the same bus line. From this, she induces that the cases may be linked, and annotates the view to share with others.

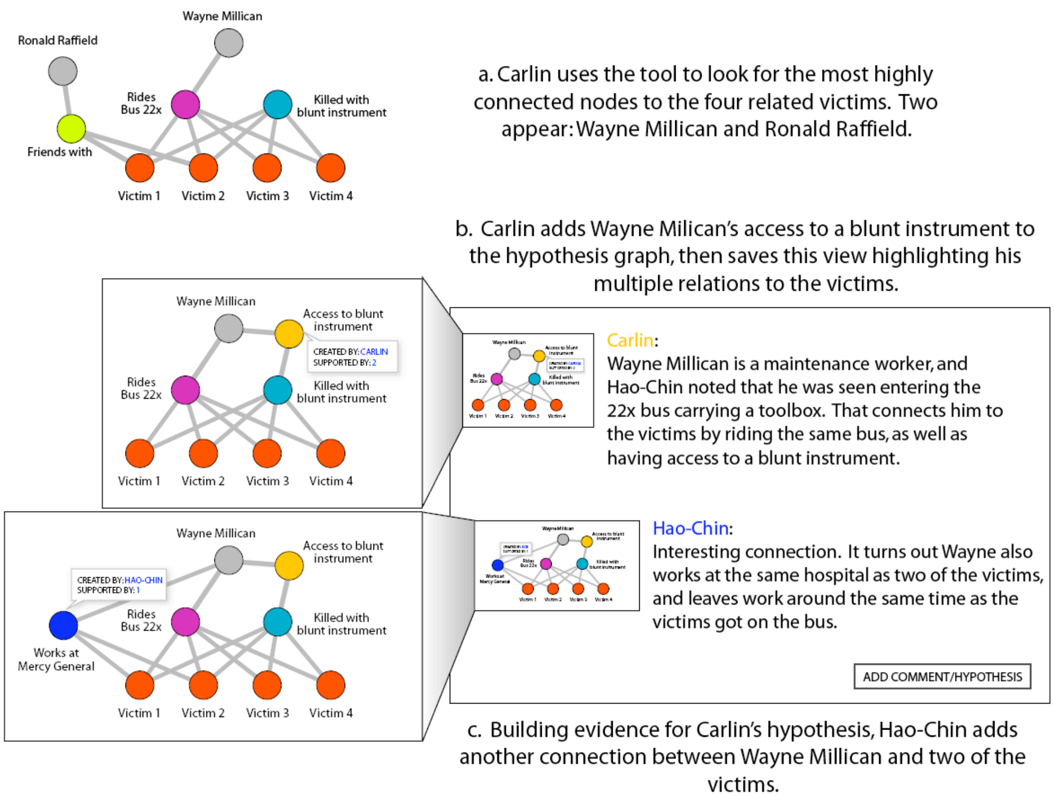

Figure 2. Generating and sharing hypotheses. (a) Carlin uses the subset of related victims to find the most highly connected potential suspects. (b) After a tip from Hao-Chin, she realizes that Wayne Millican is related in multiple ways to the victims, and annotates, saves, and shares the hypothesis. (c) Hao-Chin views the hypothesis, adds support to it, and responds to Carlin.

the bus when interviewed for an unconnected case. Realizing that Millican may be connected in more ways than currently shown, she adds a hypothesis node linking Millican to the "killed by blunt instrument" node, greatly increasing his likelihood of involvement as he is now linked to both to the bus and to the murder weapon (see Figure 2b). She then messages Hao-Chin to take a look at the hypothesis, as he may have information to add. After inspecting the hypothesis, Hao-Chin adds to it by noting that Wayne Millican has an additional connection to the victims through working at the same hospital as two of them. This updated hypothesis is saved in the same thread as the original hypothesis, so people can discuss them together (see Figure 2c).

Research questions for collaborative visualization

This scenario, though purposefully simplified, highlights some key research questions we will examine in the development of visualizations: How can categorization structures be usefully shared between analysts? How can hypotheses be created, annotated, and shared? How can surfacing traces of analyst activity benefit others? Although the most effective way to support collaboration with visualization tools will vary depending on the specific analysis task, these research questions are relevant to some of the most common tasks that analysts engage in during investigative analysis, filtering and organizing information, and building and testing hypotheses (Pirolli & Card, 2005). For all stages of this work, an iterative design and evaluation process will be used to ensure that we have solved these problems in an effective way for analysis.

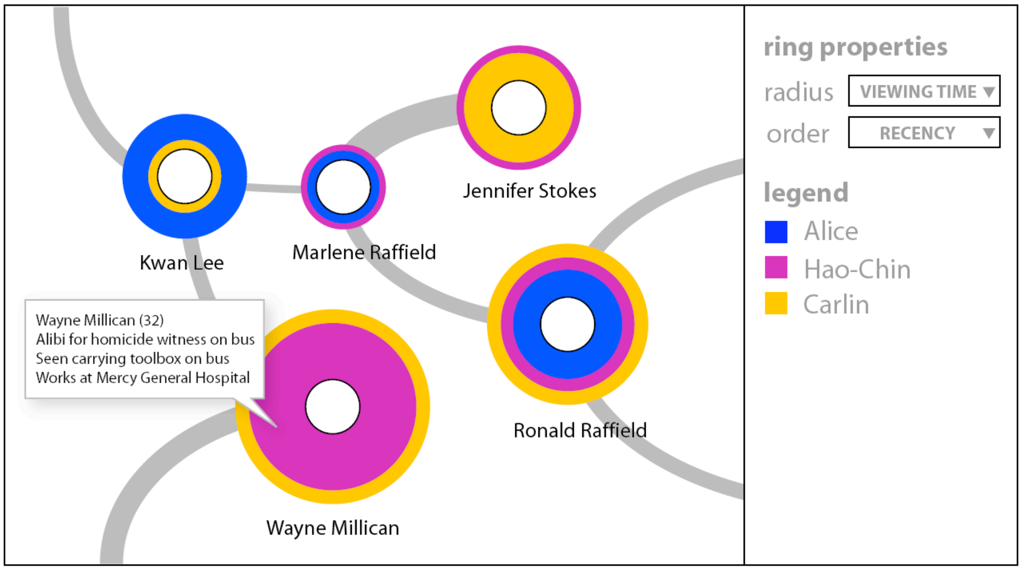

Figure 3. Surfacing analysts' traces. Investigating Wayne Millican, Carlin sees that Hao-Chin has already spent significant time examining and annotating him.

Sharing categories. Analysts may use many different ways of categorizing data 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. Each category may serve as a potential component of a nascent hypothesis, indicating a factor that might contribute to the analysts' understanding. Categories serve as a concrete, flexible, and integrable unit supporting higher-level sensemaking of the data.

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 theory-driven rather than determined by static features of the data (Barsalou, 1983; Murphy & Medin, 1985; Wittgenstein, 1953). Thus, there is no top-down correct way to categorize data; analysts will need to organize and reorganize the data in many different ways.

We will explore how to most effectively enable analysts to flexibly organize information and share those organizations with each other, using our prior work on machine-assisted organization and visualization tools as a foundation (Kittur et al., 2007; Suh et al., 2007). To support analysis in a scalable fashion, we will explore methods for creating aggregated representations of category structures and usage traces. For instance, fuzzy categorizations might be inferred from a collection of individual categorization structures, with points of overlap and disagreement automatically highlighted for investigation. An important research question to answer is whether this approach can avoid cognitive tunneling, that is, collaborators fixating on a simplified structure and ignoring other ways of structuring the data. We believe post-hoc aggregation and visual analysis of category structures constructed by people working independently may provide a mechanism to help avoid groupthink.

Our research must also tackle the perceptual scalability of visual representations of both raw data and metadata such as category structures. For instance, large networks with dense connectivity may be better depicted by applying appropriate network clustering or aggregation techniques (e.g., van Ham & van Wijk, 2004; Wattenberg, 2006) and by using a matrix display rather than a node-link diagram (Ghoniem et al., 2004). Our ongoing research on visualization architectures (Heer et al., 2005; Heer & Agrawala, 2006; Bostock & Heer, 2009) and graphical perception (Heer & Robertson, 2007; Heer et al., 2009) provides extensive experience and tools for developing scalable visualization designs.

Sharing hypotheses. 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 (Gentner, 1983; Hummel & Holyoak, 2003). Groups can promote such integration when analysts generate hypotheses that others can cycle back and 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 question here is when such schematizing produces commitment to a mental model and cognitive tunneling. The tool should allow inconsistencies in the evidence to be viewed, so that apparently similar hypotheses can be contrasted. A diversity of perspectives in a collaboration can help prevent cognitive tunneling (Convertino, et al., 2008; Jehn et al., 1999; Mohammed & Dumville, 2001). We will explore how to support this diversity, for example by making it possible to task an analyst to hunt for contradictory evidence for popular hypotheses.

Another important issue is that each analyst may have his or her own criteria for judging what information is important and relevant, and different thresholds for sharing information he or she thought important. Thus this stage would likely be sensitive to social pressures such as the cost of sharing incorrect information or trusting others enough to give them priorities before all evidence is sifted. We will examine how to promote effective hypothesis sharing based on our prior work in schema induction (Kittur et al. 2004; Kittur et al., 2006), and the literature in the social costs of information sharing (e.g., Heuer, 1999; Johnston, 2005).

Surfacing traces. With very large numbers of analysts and large amounts of data, many analysts will not be online at the same time and multiple passes through the data may be necessary. The lack of immediate communication and feedback creates an increased need for implicit coordination mechanisms (Rouse et al., 1992). For example, in large scale, asynchronous production systems such as Wikipedia, explicit coordination mechanisms such as communicating requests, do not scale up to large group sizes. By contrast, implicit coordination mechanisms, such as concentrating editing work in a small group of editors, actually increase in their effectiveness (Kittur & Kraut, 2008). On the other hand, the asynchronous nature of the work means that a person can benefit from the efforts of all those who have previously worked on the task. People can benefit from the aggregate traces of all past users, not just those currently using the system (Hill et al., 1992). Also, the aggregate organization of traces of past users may be more beneficial with more users (Furnas et al., 1987), and from having more hypotheses and annotations, than when there are fewer contributors. Our goal is to harness the traces and efforts of past contributors to benefit future contributors and to determine how collaboration visualization can improve this process.

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. Many researchers have examined surfacing traces of how others have used or contributed to an artifact (Eick et al., 1992; Hill et al., 1992; Viegas et al., 2004). A few have looked at using traces of others' activity in a group setting (Gutwin, 2002), or applied this idea to analytic tasks (Wattenberg & Kriss, 2006; Willett, Heer & Agrawala, 2007). For example, Willett et al. found that visualizing analysts' visitation patterns in navigation controls increased visitation to popular views but also led to increased inspection of unvisited views. However, little else has been done to characterize and provide guidance for the design space of such awareness cues.

We will further explore what types of information about others' analytic behavior would be useful to surface—both implicit activity such as what information analysts have looked at or searched for as well as explicit activity such as what they have discussed or marked as important. We will also examine what level of detail is most effective to surface. Depictions of activity traces might be scaled by depicting aggregates of total usage (as opposed to individuals, as in Figure 3), enabling interactive filtering to selectively show the activity of specific teams or individuals, or automatically selecting and highlighting the work of relevant analysts using activity metrics such as frequency and recency of access. Furthermore, activity traces can be used to communicate which regions of the data have been neglected, thereby helping analysts allocate their attention more effectively.

Because the visualization of traces will compete for attention with the data themselves, showing more traces may not always be beneficial. Moreover, because the traces show human behavior and not "objective" data, there is a social dimension to this tool. Analysts may be uncomfortable with others' attention on their behavior, and this discomfort may increase in organizations with social costs of false positives or when the collaboration is made up of analysts accountable to different organizations (Stewart et al., 1998). We will begin by building on our work in surfacing traces in both visualizations and Wikipedia (Willett et al, 2007; Kittur et al., 2007, 2008; Suh et al., 2007, 2008) and examine the informational and social benefits and costs of this approach. Without actual experimentation, it is not obvious what the effects of these methods will be. We therefore plan to tightly interweave the design and development of our analysis tools with behavioral studies of analysts collaborating using these tools.

Activity 2: Behavioral studies of collaborative investigative analysis

In Activity 2, we aim to investigate how the introduction of visualization tools can overcome problems that arise from scaling and to enhance our understanding of how scaling people and data influences the process of collaborative analysis using such tools. Our strategy for pursuing these goals is to look at collaborative analysis at different levels of scaling of people and data. We will begin our research in a domain where we have a developed paradigm for studying analysis—criminal analysis. Inducing useful schemas from that information is a difficult task, as there is a huge amount of information, people may specialize in different areas, and not all links are important. We also use a dataset from the SEMVAST project (www.cs.umd.edu/hcil/semvast).

In our prior experiments using the serial killer task, the total amount of data did not exceed the capabilities of a single analyst. A strategy of pooling unshared information might be effective with these limited data (e.g., Stewart, Billings, and Stasser, 1998), but not when the amount of information exceeds what a single analyst can review. To conduct the proposed research, a modification we will make to this paradigm is to add more cases, more details about each case, and more statistical data, bringing the task more in line with real-world analytic tasks. This modification will make the task more comparable to that of intelligence analysis, detection of corporate crime, and epidemiology.

Below, we refer to this task as the *Scaled Serial Killer Task.* For all studies, in addition to the college student participant pool we also plan to use either actual analysts or graduate students trained in analytical fields such as epidemiology (Cornell) or business analysis (Cornell, Carnegie Mellon, Stanford). These fields also may attract people high in analysis abilities (Frederick, 2005).

Study 1: Effects of scaling people and data

We will examine how groups of analysts exchange information and develop solutions as both the size of the group and amount of data changes. Groups of 3, 6, 9 and 12 participants will work on the *Scaled Serial Killer Task*. Each group will be responsible for either 300, 600, 900 or 1200 pieces of data. Members of the group will be assigned to solve detective cases. Half of the groups will use our previously developed social network diagram tool (Balakrishnan et al., 2008) as they work on their tasks. The other half of the groups will receive no visualization. The dependent measures will focus on how groups exchange information and whether or not they identify a serial killer embedded in the dataset.

Using this paradigm, we can examine the effects of scaling both people and data independently. We anticipate that the value of adding analysts to the group will depend on whether or not that additional analyst brings with him/her more data for the group to consider. When the analyst does not, coordination load is increased but the likelihood of finding the serial killer may not improve. In addition, we will record and code group interactions as well as individual manipulations of the data. Our prediction is that the group's strategy for analyzing data will change as information load increases. Specifically, for groups with the visualization tool, we expect there to be a shift from explicit (verbal) sharing of information to implicit (via changes to the network diagram) sharing of hypotheses as information load increases.

In addition, we will use Likert scales to assess people's view of their partners (e.g., are they doing their share? Would they work with them again? Are they competent?), of the task (e.g., How difficult is it?) and of team performance. We will measure cognitive tunneling as the overlap in categories considered by the team. We also will score performance on the task based on the number of references to evidence of the serial killer people list in their reports. In addition, we will log and analyze people's category creation processes to see when and how often they make new categories, when and how often they consider others' categories, and whether people's new categories build upon one another's categories.

Study 2: Scaling across organizational boundaries

We will examine how problems of scaling are compounded when analysts work across organizational boundaries. Groups of 3, 6, 9 or 12 analysts will try to identify the serial killer based on 1200 pieces of information. Half of the analysts will be located at Carnegie Mellon University and told that they are part of the Pennsylvania office; the other half will be located at Cornell University and told that they are part of the New York office. Half of the groups will share a single social network visualization tool; the other half will use two visualizations, one for each site ("organization"). We anticipate that a single social network tool will increase sharing of information across sites.

In addition to the measures above, we will examine how organizational boundaries affect the sharing of information, time spent considering information shared by others, and the likelihood of a solution. When initial visualization tools from Activity 1 are available, we will incorporate them into our manipulations. This study also may incorporate explicit manipulations of the social costs of information sharing, e.g., participants docked for incorrect hypotheses.

Study 3: Sharing categorization structures.

This study follows the iterative development of tools that promote the sharing of categorization structures (cf. Figure 1). We hypothesize these tools will enhance collaborative analysis, especially as group sizes become large and synchronous discussion becomes more problematic. We will use the *Scaled Serial Killer Task*. For this type of problem, where the data are filtered at the local level as relevant and organized into implicit or explicit categories (e.g., weapons, alibis, transport), visualizations that promote shared categorization may be particularly useful, although when more people are creating many categories at the same time, their utility may be decreased. In this study, groups of 3 participants will work together on the serial killer task. Half of the groups will be provided with a shared categorization structure visualization tool, the precise details of which depend upon the outcome of efforts in research activity 1. The other half of the groups will be provided with no visualization tool. Cross-cutting this manipulation will be a comparison of different lengths of initial individual exploration of the data prior to group discussion. The outcome of this study will influence subsequent studies as we scale up to more participants and new visualization techniques.

Study 4: Integrating and sharing hypotheses.

Study 4 will test the proposition that visualizations that support the sharing of hypotheses will provide value to collaborative analysis above and beyond visualizations that support sharing of categorizations. The conditions of the study will be: no visualization tool, shared categorization tool used in Study 1, shared categorization tool + shared hypothesis tool. As before, the precise features of this tool will depend upon the outcome of visualization efforts. Groups of 3 or 6 participants will work together on the serial killer task. Measures will be similar to those of Study 1. In addition, we will log and analyze use of the visualization tools and analyze the relationship between this usage and verbal exchange of information (from IM logs) and task performance. We also will measure the extent to which participants debate hypotheses, as the tool is developed in ways to compare and contrast different hypotheses. A follow-up study will examine the effect of social costs of information sharing on use of these visualizations.

Study 5: Surfacing traces.

This study explores the conditions under which people can benefit from the traces and efforts of past users to improve their investigative analysis performance. Our basic approach will be similar to that of earlier studies, in that individual participants will try to identify relationships in a large dataset. For this study, however, we need to move beyond the serial killer dataset to one with thousands of entities and relationships. Our plan is to draw on the VAST 2006 visualization competition dataset, which includes a conspiracy to commit financial fraud and a related epidemiological outbreak. These data represent problems for analysts that may be solved through visualization and augmentation tools. Significant effort from many parties has gone into the creation of these synthetic datasets, and they include very large amounts of information that would be difficult for any individual to analyze. They are thus ideal for our purposes. Our goal is to understand how to harness the past behavior of others most effectively to aid in investigative analysis. The results will inform the design of visualization tools and group organization strategies that maximize the effectiveness of large, asynchronous groups.

We will manipulate how many prior participants' work will be aggregated. Levels include aggregations of 0, 5, or 30 participants. We will aggregate the work of analysts in the 0-participant condition as input for the 5- and 30-participant conditions. Secondary independent variables will vary per experiment, such as the social cost of sharing information or users' expectations of the utility of the aggregate. Outcome measures will include objective criteria such as the accuracy of the solution, time to completion, use of others' work, and time spent on relevant versus irrelevant items. Measures will also include subjective criteria such as perceived understanding of the information space, liking of the tools, and self-efficacy.

Additional studies

In later years, we will intend to conduct additional studies to follow up on the results of Study 3 using the VAST dataset. We will address issues such as: Which past efforts are most useful to users, and what is the optimal representational level at which to surface them? How should results be altered based on properties of the analyst, such as motives, interests, and expertise? Do the quality and users' expectations of traces impact performance? We will also investigate how previous users' categorization efforts can be harnessed to support asynchronous work; i.e., when shared categories or hypotheses are from past contributors rather than current collaborators.

Research Timeline and Performance Goals

Our research timeline is shown in Table 1. Entries refer to the years that the studies are conducted; presentations and submissions for publication will follow in the next year.

Activity	**Year of Project**		
	Year 1	***Year 2***	***Year 3***
Visualization Development	*Sharing categorization structures*	*Integrating and sharing hypotheses*	*Surfacing traces*
Behavioral Studies	*Expand serial killer task* *Studies 1 and 2: Effects of scaling people and data, effects of organizational differences.*	*Adapt VAST dataset* *Studies 3 and 4: Value of sharing categories, and value of sharing hypotheses*	*Study 5: Value of sharing behavioral traces* *Follow up studies*

Table 1. Tentative schedule of research activities by funding year.

Synergistic and Educational Activities

Synergistic activities. The proposed project has close ties to other NSF-funded work at Carnegie Mellon and Cornell, including studies of coordination among hospital teams (Kiesler, Fussell), studies of multitasking and attention, studies of cultural diversity in collaboration (Fussell), studies of surfacing traces in collaboratively-generated knowledge systems (Kittur and Kraut) and studies of visualizing categorization structures (Kittur and Chau). The project also has close ties to visual analytics research at Stanford (Heer), which houses a DHS-funded Regional Visual Analytics Center. The PIs also have relationships with (unclassified) research in the intelligence community, through the KDD program, and talk with analysts through that program. Heer also has fostered a relationship with a company developing analytic tools (see attached letter).

Integrating research and education. The PIs will integrate their research with ongoing educational activities at their respective sites. Graduate students engaged in these projects interact in research team meetings, providing them with a stimulating, multidisciplinary educational environment. Kiesler will introduce a collaboration tool in her Human Factors undergraduate/masters course to help students use fault tree analysis. Fault tree analysis is a difficult skill and students commonly overlook evidence and use overly simple categorization schemas. Building tools for comparing and using others' fault tree analyses will require an algorithm that could build on commonalities in how students label levels of the tree and linkages.

Integrating diversity. The PIs have a history of integrating foreign students, minorities, and women in their research projects. Drs. Kiesler and Fussell have advised doctoral, master's and undergraduate students, of whom approximately half have been women and 75% have been minority and foreign students. Drs. Heer and Kittur will be new faculty investigators, injecting innovativeness into our work (Guimera et al., 2005).

Results of Prior NSF Funding

Large scale collaboration in critical environments. Kiesler, S., Fussell, S. R., Yang, J., Weisband, S., Xiao, Y., et al. (Sept. 1, 2003-Aug 30, 2010) (NSF ITR 0325087, with University of Maryland Medical Center and University of Arizona). Total budget: $1,250,000.

The goals of this project are to understand how people allocate time and effort across multiple projects with different collaborators and to develop new technologies to enhance coordination across people and projects. We conducted field studies on how medical personnel coordinate in hospital OR suites (Xiao et al., 2007; Ren et al., 2008; Scupelli et al., under review) and a series of lab studies examining how people allocate communication and effort across two projects, each with a different partner (Fussell et al., 2004). We also examined how cultural factors impact communication and collaboration (Setlock et al., 2007, Kayan et al., 2006, Setlock et al., 2004). Tool development efforts include a new IM tool providing information on partners' task related activities (Scupelli et al., 2005) and a video system to help remote experts allocate attention across multiple novices (Ou et al., 2005). This project has been highly successful in generating publications, conference presentations, and invited addresses, as well as in training students. To date, it has funded 4 doctoral students, one post-doctoral fellow, four Master's students, and many undergraduates through NSF's REU program. Several of the Master's and undergraduate students have gone on to industry or further study.

ITR Research Assessment. J. Cummings & S. Kiesler. Collaborative Research. (June 1/2004- June 1, 2008). (Research portion of this grant, $150,000.00)

This evaluation study focused on the coordination activities and project outcomes of 491 of NSF's Information Technology Research collaborations. The PIs ran a meeting of ITR PIs and studied their projects. An important finding was that more universities involved in an NSF-sponsored collaboration predicted fewer coordination activities and fewer project outcomes. Mediation analysis showed that insufficient coordination explained a negative relationship between the number of universities and project outcomes. Further coordination activities declined most when both number of disciplines and number of universities were high. This project informs the current proposal because it identifies coordination costs as a primary reason why scaling collaborations is a problem. In addition to a report to NSF, key publications from this project are Cummings and Kiesler (2005, 2007, 2008).

NSF Graduate Research Fellowship, A. Kittur (June 1, 2003-June 1, 2006).

This fellowship supported a series of studies examining how humans learn, represent, and use abstract relational concepts. While most previous studies of categorization have focused on concepts defined by simple features, some of the most important types of concepts–such as barrier, conservation of energy, or breach of contract–cannot be so described. Drawing on theories from analogy and schema induction, we conducted a series of experiments that showed fundamental differences between how feature-based and "relational" categories are learned, represented, and used (Kittur et al., 2004, 2006). Bayesian statistical modeling was used to quantitatively capture these differences and to demonstrate the computational constraints needed to model them (Kittur et al., 2006). Our results support a view of human categorization that is remarkably flexible, dynamic, and driven by individuals' goals rather than surface features of the environment.

Collaboration Plan

PI roles and responsibilities

The project will be directed by Dr. Sara Kiesler, in the Human Computer Interaction Institute at Carnegie Mellon University, who will be responsible for scheduling team meetings, submitting NSF reports, and other managerial activities. Dr. Kiesler will also head the efforts to conduct Studies 3 and 4, incorporating the results of our visualization development efforts. A doctoral student with skills in behavioral research and group collaborative will assist in developing the Serial Killer task and designing and running experiments. Dr. Kiesler and her doctoral student will be assisted in these activities by undergraduates interested in the area of human-computer interaction (HCI) and computer-supported collaborative work (CSCW).

Dr. Susan Fussell, in the Department of Communication and Field of Information Science at Cornell University, will head the efforts to expand the Serial Killer task and to develop the new research paradigm used to study the effects of scaling data and people on collaborative analysis. She will also head the efforts on Studies 1 and 2. A Cornell graduate student with expertise in computer-mediated communication and CSCW will work with her on these activities. One or more part-time undergraduate students will assist in preparing materials for studies and running participants. Although each has taken primarily responsibility for some studies under Activity 2, Drs. Kiesler and Fussell plan to collaborate closely on all of them.

Dr. Jeffrey Heer and Dr. Aniket Kittur will head the efforts to design and develop new visualization tools to improve collaborative analysis. Visualization tool development will be conducted at Stanford University under the supervision of Dr. Heer, who will bring to bear expertise on automating and designing the visualizations and generating large scale visualizations that people can understand. Dr. Kittur will head efforts developing and applying cognitive science theories and factors to visualization design, including what elements should be visualized, how the visualizations affect meaning and sensemaking, and how these factors can be measured. Again, though each has taken primary responsibility for tools under Activity 1, both plan to collaborate closely on each part. A graduate student with expertise in visualization and human-computer interaction at each of their institutions (Stanford, Carnegie Mellon) will assist in these efforts and spend a summer working at the other institution (Carnegie Mellon, Stanford) to facilitate collaboration. One or more part-time undergraduate students will assist in preparing visualizations for use in the experiments of Activity 2.

Project management across investigators, institutions and disciplines

Project management will be the responsibility of Dr. Kiesler. As a group, the team will develop research plans for the parts of the work for which they are responsible. Through the specific coordination mechanisms described in the next section, Dr. Kiesler and the co-PIs will ensure that research progresses as planned.

Project management across disciplines is facilitated by the tightly interwoven research plan, in which behavioral studies directly inform the development of visualization tools, which will in turn be evaluated in subsequent behavioral studies. The PIs have extensive experience with interdisciplinary collaboration, including collaboration with investigators from the set of disciplines represented in this proposal. Kiesler and Fussell have jointly advised three graduate students, a relationship which they have continued after Fussell moved to Cornell. Kiesler and Fussell have also collaborated on three large NSF grants in the past, all of which have been successful and made significant contributions to the field.

All PIs have a history of successful interdisciplinary projects, as evidenced by co-authorship of papers and ensuring relationships with faculty from other disciplines.

Specific coordination mechanisms

Semi-yearly site visits

The PIs and their graduate students will convene twice a year for a day and a half, alternating west and east coast. These meetings will help team members continue to get to know one another, learn in depth about research activities at each site, and brainstorm as a group about directions for the project. In addition, the PIs expect to meet in person several other times a year for conferences (e.g., at the CHI, CSCW and HRI program committee meetings and conferences).

Monthly PI meetings

The PIs will meet monthly to discuss progress and discuss next steps in each of the main activities of the grant. Meetings will be held over audio conference, using screen sharing software to share presentations.

Project meetings

Bi-weekly meetings will be held that include students and faculty at the three sites. At each meeting, one student will present his or her current research project and discuss open research issues with the entire group. The meetings will ensure that students and faculty at each location are actively engaged in each other's research. Meetings will generally be held over audio conferencing with additional tools to allow students to present slides and other materials via the web. On occasion, we will make use of video conferencing facilities already present at each site.

Informal communication

All PIs regularly use Instant Messaging and are accustomed to talking to one another on a near daily basis. These conversations will be used to resolve short term issues (e.g., purchasing decisions) and to keep one another posted on progress in between monthly PI meetings.

Budget line items supporting coordination mechanisms

The monthly PI meetings, weekly project meetings, and informal meetings require no additional budget items. The monthly PI meetings, weekly project meetings, and informal meetings require no additional budget items. The alternating semi-yearly site visits require two trips per year for four or five individuals (PIs and 2 students to Stanford and to Cornell or Carnegie Mellon). Carnegie Mellon and Cornell are in driving distance of one another, helping limit costs.