

The Effects of Collaboration and System Transparency on CIVE Usage: An Empirical Study and Model

Gloria Mark and Alfred Kobsa
University of California, Irvine

Contact details:

Gloria Mark
Department of Informatics
444 Computer Science
University of California, Irvine
Irvine, CA 92697

Phone: +1-949-824-5955
Fax: +1-949-824-1715
Email: gmark@ics.uci.edu

The Effects of Collaboration and System Transparency on CIVE Usage: An Empirical Study and Model

Gloria Mark and Alfred Kobsa

Abstract

We present an empirical study in which we investigated group vs. individual performance with collaborative information visualization environments (CIVEs), the effects of system transparency on users' performance and the effects of different collaborative settings on CIVE usage. Subjects searched for findings with CIVEs, working either alone, in a collocated dyad using a shared electronic whiteboard, or in a remote dyad using application sharing. Groups answered more questions correctly and took less time with the more transparent CIVE than groups using the less transparent CIVE. We interpret our results to mean that groups have better self-corrective abilities when the system is transparent. We present a stage model to explain the collaborative process of using CIVEs, which accounts for task type, collaborative setting, and system transparency.

1. Introduction

Information visualization tasks are increasingly being performed in collaborative contexts. The analysis of census or scientific data are examples of tasks that are often done by groups. The recent increase in the development of collaborative information visualization environments (CIVE) reflects the awareness of supporting group activity.

Users of CIVEs are likely to face many challenges, yet surprisingly there is a lack of studies on how groups use information visualization systems. User studies are critical for several reasons. First, they can aid in the design of CIVEs by revealing how functionality works in actual usage. This can reveal latent requirements that may not be known to designers. Second, they can show insight into how individuals or groups interact with visual data and what kinds of coordination problems they experience. For example, it can reveal the types of coordination protocols that groups develop. Third, user studies enable theories to be constructed on how people utilize visual representations of data to make decisions. Results of user studies can inform the design of systems.

In this paper we report on an empirical study of CIVEs. The goals of the study were: 1) to understand how collaborative system use may differ from individual system use, 2) to understand how different CIVEs impact collaboration, and 3) to understand how the type of collaborative setting affects CIVE usage. Our intent is to not only provide quantitative measures of performance but also to identify the processes of collaborative information visualization behavior. In this paper we expand on studies reported earlier (Mark et al., 2003; Mark et al., 2002).

We chose several variables to investigate in our study. Our motivation for choosing these variables was to address important issues concerning common usage with CIVEs. First, the differences that exist between group and individual CIVE usage are currently still unclear. We asked the question, how do groups and individuals differ in their accuracy and speed of solution in using an information visualization system? Does result quality differ between groups and individuals?

Second, organizations are becoming more collaborative and distributed (cf Castells, 1996). Current collaborative practice in organizations involves collaboration not only when people are collocated but also increasingly more when they are at a distance. Distinctions among collocated and remote collaborations are made in the field of CSCW (Olson & Olson, 2000; Johansen, 1988). We therefore investigated how these two different kinds of collaborative settings affect decision-making processes using CIVEs.

Third, in order to be able to generalize our results more broadly, we chose to investigate two different systems. We chose the dimension of transparency as a way to distinguish the systems. System transparency refers to when a user can immediately understand what the system is doing and how, by viewing the interface. The value of offering system transparency has been shown to provide benefits such as greater visibility for interactivity and feedback (Koenemann & Belkin, 1996), satisfaction (Muramatsu & Pratt, 2001), and in general, an improved understanding of the relationships between system input and output (Johnson & Johnson, 1993). We therefore investigated whether group and individual information visualization behavior differs when systems offer different degrees of transparency.

1.1 Collaborative and individual information visualization

One of the main differences between group and individual problem-solving is that group work involves coordination. Group members need to make various decisions at all phases of a collaborative decision-making process: they must decide who has what expertise, must select the decision strategy to use, negotiate roles, and must agree when a solution is reached. All of these activities involve coordination. Coordination can either be supported by a CIVE, or if not, then users must agree on their own social protocols.

Some evidence suggests that problem-solving based on visual information might be performed differently by groups and individuals. First, some studies show that the quality or accuracy of a problem solution depends on the appropriateness of the external problem representation (Anderson, 1980; Newell, 1980; Schwartz, 1971). One of the first steps in problem-solving is to choose an appropriate way to represent the information. For example, with Duncker's classic problem of how to irradiate a tumor in the body without harming the surrounding tissue, a drawing of the sun with rays spreading out around the perimeter enables people to reason through analogy that radiation can be sent to the tumor from all directions (Gick & Holyoak, 1983). Thus, the likelihood of solving the problem is higher when the information is arranged in a visual representation that is appropriate to the problem.

Individual differences in problem-solving are well-known (e.g. Norman, 1976) and in particular, individual differences in the representation of problems, displays, and data have been well-documented (e.g. Borgman, 1987). The impact on group problem-solving is important to consider: when people are in a group, a problem representation may be appropriate for one person, but not for another. Several scenarios are imaginable: one person's representation may dominate and be adopted by the group, the group may remain in conflict, or the group may successfully negotiate a common representation from individual proposals.

Other factors also affect the group process. A group may be limited by its slowest member resulting in a slower decision time than individuals. Groups are also susceptible to domination by a member, irrespective of expertise. The group process can also interact with problem type. "Eureka problems" are known to be solvable by the group in the time

that the quickest group member finds a solution (Lorge & Solomon, 1955). Groups take longer than individuals to solve problems that are more subjective in nature, such as choosing among job candidates, due to negotiation (Sproull & Kiesler, 1991).

1.2 Overview of systems used

In this experiment, we have chosen two different visualization systems: *InfoZoom* (formerly Focus; Spenke et al., 1996; Spenke & Beiken, 1999) and *Spotfire* (formerly IVEE; Ahlberg & Wistrand, 1995). InfoZoom presents data in three different views. The *wide view* shows the current data set in a table format, with rows representing the attributes and columns the objects. In the *overview mode*, the values in the rows become detached from their objects. Rows here represent the value distributions of attributes in ascending or descending order, and are independent of each other. Fig. 1 shows an example of InfoZoom's overview mode, with data from a web-based dating service containing users' self-descriptions. Under the "Gender" attribute one can see the ratio between males and females in the database, and under "Weight", the weight distribution.

Figure 1 here

InfoZoom's *compressed view*, finally, compresses the visualization horizontally to fit the window width. A row may be sorted in ascending or descending order, with the values in the other rows being rearranged accordingly to make each column represent one object. This operation reveals dependencies between attributes (like positive or negative correlations). Hierarchical sorting of two or more attributes is possible as well and can, e.g., reveal differences in the distribution of numeric attributes dependent on one or more

categorical attributes. An example is given in Fig. 2. The weights of the people in the database are hierarchically sorted in ascending order by their response to the question “Do you participate in sports?” The top line shows the weight distributions of these groups (there is a barely visible third group that did not answer this question, on the left-hand side of the ‘no’ group). The bottom line shows the mean weight per group. One can see that those who engage in sports are on average heavier than those who do not.

Figure 2 here

InfoZoom's central operation is “zooming” into information subspaces by double-clicking on attribute values, or sets/ranges of values. InfoZoom thereupon shows records only that contain the specific attribute value(s). In all three views, values of (identical adjacent) attributes become textually, numerically or symbolically displayed whenever space permits this, which facilitates the comprehensibility of the data (Kobsa, 2001).

Spotfire offers several types of mostly familiar visualizations, including scatterplots, bar charts, pie charts, graphs, parallel coordinates, trellises, etc. (in our experiments, we only used the first three types which are sufficient for our kinds of data). Unlike in InfoZoom, these different Spotfire visualizations are independent stand-alone visualizations. Two variables can be selected for display in the x and y coordinates, and a few additional variables can be selected for coding by color, shape, etc.

Fig. 3 shows the same result as the lower part Fig. 2, namely a bar chart visualization of the average weights (x axis) by response to the sports question (y axis)¹. Focusing on

¹ The bar labeled "c" indicates the average weight of those people who did not answer the sports question.

information subspaces is performed by excluding or including attribute values using sliders, check-boxes and radio buttons in the so-called “query device” in the upper right corner of Fig. 3.

Figure 3 here

1.3 System transparency

A software system is generally regarded as “transparent” when it evokes an easy-to-understand system image in users (Preece et al., 2002) or “when you can look at it and immediately understand what it is doing and how” (Raymond 2003). Using his PIE interaction model, Dix et al. (2004) define the term more formally via two functions that map the system display to the system results (“result transparency”) and to the system effect (“effect transparency”, i.e. by “effect” the authors mean the system state). If such functions exist, i.e. if the system is transparent, then the interface unambiguously predicts both system states and results. Direct-manipulative interfaces are generally regarded as more transparent than menu-based interfaces and show shorter task performance times (e.g., Ulich et al. 1991). A central characteristic of direct-manipulative interfaces is the permanent visibility of all available objects at the interface, and the fact that these objects can be manipulated at any given time. For information visualization systems, we consequently argue that the ready visibility of all data dimensions also contributes to system transparency since this enables users to construct more easily a model of the visualized data and the underlying domain.

Figure 4 here

The two systems that our subjects used seem to rank differently in terms of system transparency. In Spotfire, a significant portion of the system's functionality is not immediately visible and directly accessible. Instead, a "Properties" menu must be accessed which contains controls for numerous parameters that can influence the shape of the visualization (see Fig. 4). On the data visualization side, Spotfire displays relatively few variables that must be deliberately selected. Kobsa (2001) found that Spotfire therefore imposes fairly high "cognitive setup costs" on users: they needed considerable time to choose among the different visualizations that Spotfire offers and to set them up correctly, particularly when solutions included several variables and therefore several steps. Users must plan ahead what variables to use and how to visualize them. Selecting a visualization for the first one or two variables imposes severe restrictions on the display of the remaining variables. The upfront planning must therefore be very thorough and comprehensive so as to avoid dead ends. This planning moreover must be performed without assistance from a visualization and takes considerable time.

Users of InfoZoom, in contrast, can rely much more often on visual cues when accessing both data and system functionality. (The only major exceptions are the dialog windows for defining new "derived" variables and for the charting function, both of which violate the otherwise relatively straightforward "click on what you want" paradigm.) Kobsa (2001) found that users can interact quite effortlessly with the system. Users can also plan incrementally, i.e. perform a few steps, see how far they have come,

and proceed or switch to a different view if the partial plan turned out to be wrong (some users even developed a "click first - think afterwards" problem solving behavior).

1.4 Related work

One of the main challenges for CIVEs is to support user synchronization and coordination. CIVEs can be basically categorized into the amount of coordination support that they offer users.

Systems that enable all users to fully control the data must rely on users to develop social protocols to govern their coordination. Users must together navigate through the data, filter it, choose parameters, views, and output. CIVEs of this type include a prototype for geographical data visualization (Brewer et al., (2000), and the Shastra environment (Anupam and Bajaj, 1993). Similar to the remote condition in our study, CoVis (Edelson et al. 1996) uses application sharing and desktop teleconferencing to enable groups to collaborate on the same data. Thus, users need to self-coordinate, for example in sharing control of the system.

Systems that restrict shared control of data provide a tradeoff with the coordination problem: it helps users coordinate by limiting their actions, yet users lack some flexibility to manipulate data. Such systems include HIGHEND MVE (Pagendarm and Walter, 1993), the PAGEIN project (Grave et al., 1994), and MICE (Bourne et al., 1998). DIVA (Schönhage, 2000) employs interaction protocols to coordinate multiple visualizations.

Some CIVEs have begun to explore the benefits of providing awareness information to help users coordinate. Comotion (Chuah and Roth, 2003) supplies meta-information about users' actions, e.g. their level of participation and history of interaction with objects. Cspray (Pang & Wittenbrink, 1997) enables users to mark parts of the

visualization for awareness. Several systems address the coordination problem by providing both public and private views of the data. A CIVE designed for visualizing fluid dynamics provides both types of views of the data (Gerald-Yamasaki, 1993). Similarly, COVISA (Wood, 1998) allows users to view their own data but share certain data. A prototype for GroupSystems also allows annotations as well as public views and individual nodes (Chen et al., 1998). VR-VIBE (Snowden and Jaa-Aro, 1997) is an unusual hybrid in that users can create their own subjective views of the data while still maintaining awareness of the overall process. Another method for aiding coordination is to integrate existing communication devices with the system, such as walkie-talkies, as used in a system by Rogers et al., (2002) to support nomadic workers.

Personalized avatars and audio are used in the CVD and Cave6D systems (Lascara et al, 1999) and TIDE (Sawant et al., 2000), and iScape (Börner 2001) to convey the presence of remote users by showing their locations and actions. These systems are designed to support an immersive virtual experience with the data.

No other studies have directly examined the effect of collaborative settings and different systems on CIVE usage, but we review some related work. Single user studies of information visualization systems have been conducted (Blockeel & Moyle, 2002; González & Kobsa, 2003). Park et al., (2000) studied behavior in a collaborative virtual environment. Tromp et al., (2003) assessed group behavior using a virtual reality system. Case studies of single-user innovative virtual reality tools or task-specific visualization tools include those for business processes (e.g. Schönhage, 2000) and code visualization (Jerding et al., 1997). In the field of Computer-Supported Cooperative Work, a large number of collaborative systems have been evaluated, with examples including electronic

calendars (Pirolli & Card, 1995; Palen & Grudin, 1995) electronic shared workspaces (Mark, 2002), video (Isaacs & Tang, 1993), and collaborative virtual environments (Bowers et al., 1996).

No studies to our knowledge have contrasted collocated collaboration using an electronic whiteboard with remote application sharing, as we do in our study comparing collocated and remote CIVE usage. Other technologies that can support collocated collaboration include CoLab at Xerox PARC (Stefik et al., 1987) and the CaptureLab (Halonen et al., 1990), which were early efforts that introduced collaborative technologies into the conference room for supporting collocated work. Initiatives such as Whiteboards (Donahue and Widom, 1986) Liveboard (Elrod et al., 1992), Tivoli (Pedersen et al., 1993), Flatland (Mynatt et al., 1999), BlueBoard (Russell et al., 2002) and the Plasma poster Network (Churchill et al., 2003) have expanded the collaborative potential of electronic whiteboards. Other initiatives have investigated application sharing for both collocated and remote work, such as VideoDraw (Tang, 1990), Montage (Tang et al., 1994), Cognoter (Tatar et al., 1991) as well as the evaluation of shared single-user applications (Greenberg, 1990). Li and Li (2002) evaluated groupware system adoption in terms of transparency, which relates to our discussion of transparency. Bradner and Mark (2001) studied videoconferencing and application sharing on task performance.

2. The Experiment

We chose to perform an experiment, as it would best enable us to examine differences between groups and individuals, system transparency, and the effect of collaborative setting.

2.1 Method

One hundred undergraduate students with majors in Information and Computer Science or Engineering at the University of California, Irvine participated in the experiment in the spring of 2001. Subjects qualified for the experiment if they had at least one year of computer usage. They received \$25 for their participation and competed for a \$100 prize for the best results in the discovery task of the experiment.

2.1.1 Experimental Tasks.

There are different purposes for using information visualization systems. In a survey of daily activities of administrative data analysts, González and Kobsa (2003) found that answering focused questions is currently the most frequently occurring type of data analysis. Such tasks are when one has a very specific query in the data. For example, with census data, one might ask, “What proportion of the people in the U.S. are Catholic?” With focused questions correct answers can be measured. A second type of task is when people use information visualization systems for exploratory analysis. An example of such a free discovery task is when a researcher tries to detect differences between segments of the U.S. population using census data. Open-ended free discovery tasks enable us to measure how many findings people can retrieve from data, as well as to examine the quality of their findings. We chose these two distinct types of tasks to test as we felt they simulated real world tasks, and we expected that the group coordination required by these task types might be different.

In the first task, subjects used anonymized data from an online dating service to answer ten specific questions, such as: "Did males cheat more on their girlfriends than females on their boyfriends?" This problem resembles a “Eureka problem” (Lorge and

Solomon, 1955) in that the problem has only one correct answer that is immediately recognizable once it is located on the information display. Subjects answered ten focused questions.

In the second task, subjects were given 40 minutes and instructed to discover as many findings in the data of a population survey as they could. This task is free-form, and no specific background knowledge is required to comprehend the data. Subjects needed to make decisions as to whether a finding was relevant, unlike the first task where only one correct answer exists and is generally recognized as correct once it is seen.

2.1.2 Experimental Design.

The experiment used a two-factor between-subjects design. The main factors were:

1) *System - two factor levels:*

a) *InfoZoom:* subjects used the InfoZoom system in both tasks.

b) *Spotfire:* subjects used the Spotfire system in both tasks.

2) *Interaction type - three factor levels:*

a) *Alone:* subjects worked on the task themselves at a workstation.

b) *Remote (groups):* two subjects worked together and sat at workstations in adjacent rooms. They performed the task while interacting via Microsoft NetMeeting and a speakerphone. They used either InfoZoom or Spotfire as shared applications. Subjects did not see each other.

c) *Collocated (groups):* two subjects worked side-by-side in front of a large 60” diagonal touch-sensitive electronic Whiteboard (Smart-Board), using either InfoZoom or Spotfire.

Table 1: Number of subjects assigned per condition

Condition	InfoZoom	Spotfire	Total Subjects
Alone	10	10	20
Remote	10 groups x 2 = 20	10 groups x 2 = 20	40
Collocated	10 groups x 2 = 20	10 groups x 2 = 20	40
Total	50	50	100

Subjects were randomly assigned to one of the combined conditions of System and Interaction Type. In the Remote and Collocated conditions, subjects were unacquainted with each other. Table 1 shows the number of subjects assigned per condition.

2.1.3 Measures.

In the focused question task, we measured the correctness of responses based on subjects' written answers on paper. For the open-ended discovery task, we measured the number, accuracy, and significance of findings, also based on written descriptions.

All subjects were audio and videotaped. We also recorded their screen activity using a screen capture utility. The Alone and Remote subjects were each videotaped using an IBM computer camera mounted on the top of their computer. The Collocated subjects were videotaped using a regular video camera set up approximately 5 feet from the Smartboard. In neither condition could subjects view their images on the video.

Two different coders analyzed 75 hours of videotapes. The screen recordings and videotapes were viewed jointly. Grounded theory (Strauss and Corbin, 1998), a research methodology that involves the systematic generation of theory from data, was used to identify concepts unique and common to the conditions.

2.1.4 Procedure.

In all conditions, subjects first received a 30-minute training on the visualization system they were assigned. During the first 10 minutes of the training, subjects received a general demonstration of the main system functionality, followed by a 20-minute hands-on tutorial using six questions from an auto statistics database. Experimenters verified that subjects covered all questions and, if necessary, helped answer additional questions. After the training, subjects were familiarized with the content of the two datasets to be used in the tasks. Subjects in the Remote and Collocated conditions additionally practiced using their collaborative systems of Netmeeting and the electronic whiteboard.

The first experimental task (focused questions) took 30 minutes and the subsequent free discovery task took 40 minutes. Subjects then filled in a short questionnaire and were interviewed about their experience with the system.

3. Results

Observations of the videotapes revealed that subjects had little difficulty using the systems to perform the tasks. This was also confirmed by their responses in the final questionnaire. They also easily adjusted to the usage of NetMeeting and the Smartboard. We present the results for the focused question task first.

3.1 Focused questions.

For the focused questions, coders counted the correct number of responses out of ten questions, based on verifying correctness beforehand. Table 2 shows the mean number of correct responses, by condition.

Table 2: Mean and standard deviation of correct responses to ten questions in the first task.

System Interaction	InfoZoom	Spotfire	Average
Alone	6.8 (1.7)	6.7 (1.6)	6.8 (1.6)
Remote	8.4 (1.0)	7.2 (2.1)	7.8 (1.7)
Collocated	8.0 (2.0)	6.7 (1.2)	7.4 (1.8)
Groups	8.2 (1.6)	7.0 (1.7)	7.6 (1.7)
Combined			

An ANOVA shows that subjects in the InfoZoom condition answered more questions correctly than in the Spotfire condition, $F(1,54)=4.1$, $p<.05$. Although there was a trend towards more correct responses in both Remote and Collocated group conditions, it did not reach significance at the .05 level: $F(2, 54)=2.1$, $p<.1$. There was no significant interaction, i.e. using a particular system did not affect the Alone, Remote, and Collocated conditions differently.

As there was no difference between the two group conditions, we combined their results to compare whether being in *any* group differed from being alone when using the systems. An ANOVA shows that being in a group (see Groups Combined in Table 2), yielded higher correct responses than working alone ($F(1,56)=3.4$, $p<.07$). Broken down by system, with InfoZoom, the responses of groups were significantly more correct than those of individuals ($F(1, 28)=5.0$, $p<.03$), while no difference was found for Spotfire.

Table 3: Average completion time (and standard deviation) per problem in seconds, for focused questions

System Interaction	InfoZoom	Spotfire	Average
Alone	116(103)	123(100)	120(101)
Remote	134(101)	176(156)	154(131)
Collocated	115(114)	169(145)	141(132)
Group combined avg.	125(107)	172(150)	149(129)

Table 3 shows the average completion time per problem for the focused questions. An ANOVA indicates that subjects using InfoZoom were significantly faster in solving the problems than subjects using Spotfire: $F(1,58)=6.8$, $p<.01$. Remote and Collocated conditions showed no difference, and when all groups were combined, an ANOVA showed that groups were slower than individuals: $F(2,58)=5.6$, $p<.02$. There was no significant interaction.

Thus, for the focused question task, being in a group (whether working remotely or face-to-face) resulted in more correct responses than being alone. This difference, however, is due to using the more transparent system (InfoZoom).

3.2 Free discovery task.

For the second task, making discoveries with the data, one extreme outlier was eliminated from the Alone condition since the high number of responses of this subject were mostly nonsensical. Table 4 gives a summary of the findings. The column on the right shows the average of both Remote and Collocated conditions combined, to enable a

simple individual and group contrast. Subjects in the Remote condition made the most discoveries using InfoZoom, nearly 50% more than Alone or Collocated. However, an ANOVA showed the difference between Interaction Types did not reach significance, $F(1,53)=1.8$, $p<.17$. This was due to the large variability within each Interaction Type condition. There was a significant difference between systems, however. Subjects using InfoZoom produced significantly more findings than subjects using Spotfire, $F(1,53)=4.8$, $p<.03$. There was no significant interaction.

However, looking at the raw total of findings can be misleading. Findings can be false, and they can also be meaningless. We first verified the correctness of all findings that subjects reported, and computed the ratio of accurate findings to overall findings. An ANOVA did not reveal differences between Remote and Collocated, so we combined these conditions. Using combined groups, there was a significant Group by System interaction, $F(1, 55)=3.9$, $p<.05$. There were no main effects, i.e. effects of Interaction Type or System. The interaction means that groups were more accurate than individuals when using InfoZoom. With Spotfire, groups and individuals showed no difference in accuracy.

We next looked at the proportion of “meaningful” findings. We developed criteria of what constitutes a meaningful result. Such a result must include a comparison between variables, indicate a minimum or maximum, and/or it must have “surprise” value. For example, a result of “The average age of a male construction worker that lives in the south, and is married, is about 44 years” would not be considered a particularly meaningful finding, as this statement alone is only descriptive and not very informative. It does not contrast this result with another that would give it more meaning. On the other

hand, a result like “union members earn more than other workers” would be considered meaningful as it represents a comparison. Two coders first coded sample results to calibrate themselves. They then coded the results independently, with an agreement of 95.5%. Where there was disagreement, the coders discussed each coded result until they reached agreement. Most coder discrepancies were due to oversights. An ANOVA on the proportion of meaningful results showed that the difference in the System condition approached significance, $F(1,53)=3.5$, $p<.07$. Using Spotfire yielded more meaningful results. There was no significant effect of Interaction Type, and no interaction.

Table 4: Summary of free data discovery task. Means and standard deviations are shown.

		Alone	Remote	Collocated	Groups Comb.
Mean no. of findings	InfoZoom	14.2 (6.9)	21.10 (11.5)	13.9 (6.8)	17.5 (9.9)
	Spotfire	11.8 (4.8)	12.7 (5.8)	12.3 (5.8)	12.5 (5.6)
Correctness	InfoZoom	.75 (.05)	.87 (.05)	.87 (.05)	.87 (.13)
	Spotfire	.84 (.05)	.80 (.05)	.81 (.05)	.80 (.12)
Meaningfulness	InfoZoom	.81 (.25)	.79 (.27)	.79 (.26)	.79 (.26)
	Spotfire	.84 (.17)	.87 (.13)	.97 (.03)	.92 (.10)
Net Productivity	InfoZoom	23.1 (11.4)	28.5 (21.0)	20.5 (9.9)	24.5 (16.5)
	Spotfire	16.6 (8.5)	18.8 (6.9)	19.7 (8.6)	19.3 (7.6)

We also investigated the complexity of the findings. For instance, “Females with 30 or more years of work experience make less per hour (\$9.47) than males with 30 or more years of work experience (\$15.02)” refers to the variables of gender, years of work experience, and hourly wage. This is more complex than a simple descriptive finding of

“There are 23% working in the service sector” (one variable, namely sector). We defined complexity as the number of variables involved and found that Spotfire results (mean=1.52, sd=2.0) had significantly more findings containing one variable (i.e. complexity=1) than InfoZoom (mean=.27, sd=.52), $F(1,53)=11.8$, $p<.001$. In other words, Spotfire use resulted in findings that contained fewer variables.

We last computed a coarse measure of the total "net productivity" of each data discovery session by adding the number of correct and meaningful findings weighed by their complexities. In other words, all correct findings that were judged meaningful would each be multiplied by the number of variables each answer contained and summed up. An ANOVA showed a main effect of System that approached significance (InfoZoom mean=24.0, sd=4.4; Spotfire mean=18.4, sd=4.4), $F(1,53)=3.3$, $p<.08$, but no Interaction Type effect and no interaction.

4. A model of the process of collaborative information visualization

Though our results did not yield differences between Remote and Collocated conditions, it is important to keep in mind that we focused only on the quantitative end results. These quantitative measures do not capture differences in group processes, which may exist due to the different collaborative settings. To gain a better understanding of the group process, we felt that it would be beneficial to perform a qualitative analysis of the behavior from the videotapes.

In this section, we present a model based on a qualitative evaluation of the group behavior. We describe how the model can take into account the type of task, the kind of

tool used, and the nature of the collaborative setting. Our model is based on an analysis of the verbal interaction and system behavior in the group conditions only. In the Alone condition, we could not infer the reasoning processes, as there was no verbal interaction and the recorded screen activity did not provide enough clues to inform us of the individual subjects' processes. The model therefore only applies to collaborative settings.

4.1 The model

We discovered that the process of collaborative information visualization occurs in a series of five stages (see figure 5). Though some differences existed according to task, tool, and collaborative configuration, the model can still explain the basic processes.

Figure 5 here

In the first stage of the process for the focused question task, subjects read the question and parsed it into distinct variables. The subjects identified variables in the question that were represented in the visualization programs. For example, the question: *“Are the people who do not participate in sports heavier than the rest?”* was parsed into phrases containing variables, i.e. “participate in sports” and “heavier”.

In stage 2, subjects then mapped one variable from the question (usually the first variable that appeared) to the corresponding variable representation in the visualization system. For example, the variable, “participate in sports” was mapped onto the variable in the data set that was labeled “engage in sports?” (see figure 1 in section 1.2).

In stage 3, users manipulated the system to find the appropriate visualization so that the variable could be visualized in a meaningful way. For the variable “participate in sports”, subjects needed to compare people who participated in sports with those who did not. They therefore chose visualizations that displayed each of these representations.

In stage 4, users identified and validated the results from the variable that they had just examined, reassessing whether the results helped them progress towards the overall goal of solving the problem. If there was agreement about the appropriateness of the visualization and on a plan of how to proceed, there was little verbal interaction. However, if there was disagreement, an extensive interaction and negotiation ensued, often with a change of cursor control. In the free discovery task, validation was somewhat different: subjects made sure that they were making progress towards interesting findings. If so, they proceeded to the next variable. If not, they discussed how to rectify the situation (which was often associated with a change of cursor control between subjects). Once the validation was completed for the current variable, subjects cycled back to Stage 2 in the process, choosing another variable in the problem.

The process of trying to discover how to represent a current variable sometimes revealed incorrect reasoning about a previous variable. For example, reasoning about the variable “heavier” could reveal incorrect thinking about the previous variable “participate in sports”. In such a case, subjects would return to re-examine the variable “participate in sports” in terms of the proportions of those who participated in sports and those who did not, based on a person’s weight.

In the fifth stage, the users validated the entire answer and recorded their results. Subjects explicitly conferred with each other to make sure that their answer satisfied

every aspect of the question. This stage primarily involved verbal interaction and little system use. Subjects then translated their numerical results into written prose, recording it on paper. The main challenge for subjects at this stage was to translate the visual representation into a written representation of the results.

4.2 Explaining task types with the model

The stages in the model can account for both focused question and free discovery tasks. The main difference between these tasks is the amount of constraint involved in the search process. The variables in a focused question are more clearly identifiable which leads the search process to be fairly directed, or constrained. In contrast, free discovery tasks pose general questions and do not specify variables. The actors must choose variables and analyze their values in order to answer the general question. The search space in free discovery tasks can potentially consist of the entire dataset and actors need to bound the search space.

Whereas with the focused questions, actors had to parse the question into variables in stage 1, with the free discovery tasks, the actors had to instead choose the variables to search for. The actors' goal is to discover "interesting" results. Thus, the stage 1 activity of parsing the question becomes two substages: first choosing an interesting proposition to explore and then parsing this proposition into variables that could be searched for in the dataset. The groups usually articulated a two-variable question of interest as a starting point in stage 1 to focus on distinct variables. For example, with the population survey data one subject might say, "*Let's look at female wages.*" The subjects then parsed this proposal into "female" and "wages".

With both tasks, the subjects mapped variables to the visualization, searched for the best visualization and then validated that the results from the variable were correct. However, in the free discovery task, the subjects cycled more through stages 2 through 4 until a potential “interesting” result was found (see figure 5). Groups did not always stick to their original propositions from stage 1. They often added more variables or abandoned previous variables in order to find a result that they judged “interesting”.

Stage 5 functioned slightly differently in the free discovery task compared to the focused question task. Only when the groups judged the results to be “interesting” did validation and translation of the answer occur. As subjects did not begin with a pre-defined question, there were often multiple ways of describing the results. For instance in the example above, groups could either report the wages of union vs. non-union females or make an explicit comparison, as in:

“Females in unions earn, on average, \$10.50 per hour while females not in unions earn \$8.79” or

“Females in unions earn, on average, more than those not in unions.”

At the end of stage 5, other differences were observed between the tasks. In the focused questions, once the translation of the visual into written results was done, subjects moved on to the next question. In the free discovery tasks, subjects did not start over with new variables but instead tended to build upon previous results. Subjects thus usually began a new round of searching for variables at stage 2. Returning to stage 1 was only done if groups felt that they had exhausted all the “interesting” findings with the current proposition and must begin with a new proposition.

4.3 Explaining system differences with the model

The model also accounts for the system differences. We consider Infozoom to be a more transparent system as it provides more visual cues for accessing the data. In contrast, much of the system functionality in Spotfire is not immediately visible in the interface. System differences did not affect the activity in stage 1, where the actors parsed the question into variables. System differences were most evident in stages 2, 3, and 4 of the model due to the differences in system transparency (see the right-hand side of figure 5).

In stage 2, where the actors map the variables to the program, InfoZoom users made fewer errors. All variable names in this system are in a list on the left-hand side of the screen, and almost all are visible. In contrast, with Spotfire, relevant variables can appear in either or both of two screen spaces. Often relevant variables must be scrolled through on the top right hand section of the screen. Because all the variables are not visible on one screen, errors in mapping the variables can occur. Therefore, the way that variables are displayed on the visualization system interface have an impact on how the activity in stage 2 is conducted.

Stage 3 described the process that was done with both systems. However, a focus on stage 3 revealed that Spotfire users tended to make more errors in this part of the process than Infozoom users. Spotfire has more options with which to manipulate the information. Actors could manipulate data using the query devices or the XY chart. The actors required more time to assess the different visualizations and to choose the most appropriate one for the data. It was especially problematic when more than one variable was being considered. In contrast, InfoZoom users had only one view of the data, and

therefore they tended to have fewer problems manipulating the system to find a correct view. Thus, though Spotfire offered more choices for representations, the choices served to create a higher cognitive overload for the groups. Stage 3 is thus critical when assessing the impact of system transparency on usage.

We observed no differences in the two systems in stage 4, i.e. in validating the results of the visualization. Validation at this stage involves mostly human interaction, and little system interaction, and thus groups seemed to be able to interpret the results from InfoZoom and Spotfire equally. In stage 5, where the actors validated the entire answer, this stage also described the group interaction similarly with both systems. Even though the interfaces were different, the process of translating visual into written results was quite consistent.

4.4 Explaining collaboration setting with the model

The model also accounted well for the two different collaborative settings: collocated and remote. The social dynamics of the different collaborative settings were different (see the top part of figure 5). In the collocated and remote conditions, subjects adopted quite different roles. In the remote condition, the subjects communicated and interacted with each other to a higher degree throughout all stages of the process. They adopted distinct roles to coordinate their actions with the system across distance. In the collocated condition, verbal communication between the two was far less; one subject was the primary system user while the other observed, usually maintaining a social distance (Hall, 1966), away from the Smartboard.

With Netmeeting use, both subjects participated collaboratively throughout all stages of the process. Subjects tended to divide their labor into distinct roles of “system

operator” and “system director”. The system operator manipulated the system while the system director gave instructions to the remote partner on how to interact with the system. The system director monitored the overall progress towards the goal of answering the questions. If the system operator became “lost” and unsure of what to do with the system, cursor control was then switched to the system director.

With Smartboard use, one subject primarily worked alone through all stages until stage 5, when both subjects participated in validating the whole answer. Groups in this condition tended to divide their labor into distinct roles of “system operator/user” and “observer”. The person who stood on the left-hand side of the screen controlled the application, as most controls for manipulating the visualization were located on the left side of the screen. After a few minutes, the person who initially stood on the right hand side of the screen soon sat down and became the scribe for the pair. Negotiation of one’s position in front of the whiteboard was done during the first few questions. The system operator/user stood by the Smartboard, operated the system, and made decisions throughout all the stages of the process. The observer watched and generally only interacted during stage 5, in validating the entire answer. For both collocated and remote groups, once the role of the controller of the application was set in place, control was transferred only if the controller became “lost”.

Thus, in the remote condition, stages 1-5 of the process were done collaboratively by both members in the group: parsing the question, mapping the variable to the program, finding the best visualization, and validating the results from the current variable. In the collocated condition, stages 1-4 were done primarily by the person who was operating the system, standing by the Smartboard. In stage 5, both actors discussed and validated the

entire answer. Thus, even though two different collaborative settings were used, the process stages remained the same.

5. DISCUSSION

We found coordination differences in CIVE usage depending on the system, the task, and the collaborative setting. We chose to test users with different systems in order to be able to generalize our results to a broader range of systems. However, in the free discovery task, we found a significant Group² by System interaction in response accuracy. Using the more transparent system (InfoZoom) resulted in more correct responses from groups than the less transparent system (Spotfire). Transparency affects usage in that there is less planning for groups in finding the appropriate problem representation due to the ready visibility of all data dimensions. InfoZoom offers users a full view of all the variables, which results in fewer errors in mapping the variables to the representations in the system. Spotfire offers more options (i.e. more data views) for visualizations, which increases the coordination cost. More choices will require users longer to find the appropriate visualization. More data views can potentially lead to more errors, as users can pick the wrong data view for making decisions (e.g. a pie chart as opposed to a scatter plot). However, it is important to note that we only compared two systems. More research is needed to develop the notion of transparency and how it affects system use.

² See section 3.1: both remote and collocated group conditions were combined as we found no statistical differences between them.

5.1 Task type and coordination

Focused question tasks involve a very different type of coordination process for groups than free discovery tasks. For focused questions, subjects followed a deterministic path to the solution; there was one correct solution and clear goal. The choice of the representation (i.e. visual information display) was therefore crucial to determining whether the answer was found. Once the appropriate problem representation was found, the answer could be identified by anyone in the group, as only one correct response exists. This is similar to the superior group performance found with Eureka problems (Lorge and Solomon, 1955).

In contrast, the collaborative information visualization process with free discovery tasks is more opportunistic. The groups defined variables of interest as they interacted with the system. Therefore much coordination is needed. The choice of visualization is critical as it affects the likelihood to discover relationships, and thus it affects the quality of the results.

Coordination differences can also be observed in the overall validation process (stage 5 of our model). In a free discovery task, it is much more difficult to retrace one's steps to validate an answer against the visualization. As the process is opportunistic, users do not always remember the path they have chosen. We observed that more validation occurred with free discovery tasks compared to focused questions, which indicates the extra effort involved. The ability to retrace one's steps through the system to validate an answer is important. In the free-discovery tasks, neither system supported well the users' ability to view their history of variable exploration.

5.2 Collaborative setting and coordination

Both of the systems we used in the study allowed full shared control of the system. As discussed earlier, with such systems, users must agree on social protocols to coordinate themselves. We found that the type of collaborative setting affected the coordination strategy that subjects used.

First, groups developed more explicit social protocols when they were remote. Subjects frequently articulated their plans out loud (e.g., choosing an attribute) and waited for their remote partner's agreement before they proceeded. In contrast, subjects in the collocated setting rarely confirmed their plans explicitly. Instead, gestures were used frequently to direct the partner's attention to specific points on the whiteboard screen. In the remote setting, people used the cursor to refer to areas, but often the person controlling the cursor had to ask their partner more than once to attend to where the cursor was pointing. The position and function of the cursor could be ambiguous, e.g. it was not clear whether the cursor was at rest or actually "referring" to data. In contrast, gestures to the whiteboard were less ambiguous; it clearly indicated a reference to the data in that part of the display.

Collocated partners work within a shared context where each is aware that the information is shared. However, remote partners need to explicitly and continually establish a shared context. Thus, coordination in the remote condition requires more effort as system interactions or plans need to be explicit. Further, both partners are aware that the system director may need to take over system control at any time. By continually establishing a shared context, exchange of control can occur with as little reorienting overhead as possible.

Overall, in the remote condition we observed more active participation and engagement of both subjects compared to the collocated condition. The video analysis revealed a pattern of roles that occurred in the remote condition between system “director” and system “operator”. We interpret this division of labor as a way to efficiently distribute group resources. The system director could devote all of her attention to reasoning about the problem, while the other partner devoted attention to controlling the system. In contrast, working collocated did not lead to equal participation among the partners as we expected. The dyad members divided themselves into roles of system operator/user and observer. However, the reason for lack of equal engagement could be due to the physical size limitations of the electronic whiteboard interface which could have prevented the dyad members from working comfortably side by side. The adoption of different roles therefore served as a coordination device. Our observations suggest that interface size in collocated work affects the use of CIVEs.

We did not find differences in accuracy or time to completion between the remote and collocated conditions. One reason for this could be that in both conditions, both subjects participated to a high degree in validating the answer. Thus, with two people checking answers, there was a greater likelihood of finding errors. This suggests that the validation stage is especially critical in the design and use of CIVEs.

5.3 Comparing our results with other models

Another stage model of collaborative behavior is that of Park et al. (2000), based on an analysis of CAVE6D, an immersive virtual environment. Park et al. found stages of: 1) problem interpretation, 2) agreement of parameters to use, 3) independent search for a pertinent representation, 4) reporting the discovery to their partners, and 5) negotiating

the conclusion of their representation. The first three stages of this model are similar to our model. The main difference from our study is that dyad members in the Park et al. study worked independently. Their system provided private data views, and in stages three and four partners searched independently for data and reported results back to the other. Our model instead applies to systems where users work synchronously with the same data views. A major limitation of the Park et al. study was the small sample size, so it should rather be viewed as a pilot study.

Though our study was performed *in vitro*, we feel that the collaborative behaviors we observed were in response to the ecological constraints of the situation, i.e. adopting different roles depending on whether the work was remote or collocated. This was also evident in the free discovery task where the dyads changed their paths as new opportunities arose for finding new variables. We therefore believe that our model has some degree of ecological validity.

Another theory that also accounts for information seeking behavior in an ecological context is information foraging theory (Pirolli and Card, 1995; Chi et al., 2000; Pirolli and Card, 1999), which applies a metaphor of food foraging behavior to information seeking behavior. People seek the most and best information available with the least expenditure in energy. Though information foraging theory is designed to explain single user behavior, there are some similarities with our model of collaborative information visualization behavior. Our model also can explain that the dyad partners behaved so as to maximize their information search by reducing coordination costs. First, in both collocated and remote conditions, the groups divided up their labor, which could have maximized their efficiency. Working in parallel reduces coordination costs as people are

working on separate tasks. However, coordination is needed at the outset and for combining results once they are found. Second, in collaborative information visualization, it is efficient for groups to collectively validate the results, as more people can check and question the results and spot errors. In both collocated and remote conditions, the dyad members collectively participated in the validation stage. Information foraging theory predicts that people modify their information search according to the preliminary results they find. This is consistent with the behavior that we found for free discovery tasks. The group members adapted their search behavior in an opportunistic fashion according to the results they found.

However, there are some important differences between our model and information foraging theory. Information scent refers to the strength of cues that guide the search process. In collaborative information visualization, it would seem to be more optimal for all group members to look for cues that indicate the relevance of a particular path, i.e. in mapping variables to the program and finding the correct visualization (stages 2 and 3 of our model). Yet we found this was not the case. However, strategic planning theory (Berger, 1997; Gattis, 2002) could explain the behavior of dividing up the labor in some stages of the search process. Strategic planning theory accounts for collaborative behavior, unlike information foraging theory. People strive to achieve metagoals, which include efficiency but also social goals, such as achieving good social relations with the group or maintaining one's professional reputation. Social goals may override efficiency. Thus, it is possible that the dyads divided their labor to perform efficiently, but also because this was the most socially acceptable way in the collocated condition to collaborate, given the interface. In the remote condition, passing control back and forth

could be very awkward. It was less awkward to adopt roles where one person continually interacted with the system while the other directed. In the collocated condition, working side by side would violate the limits of social distance, which according to Hall (1966) should be at least 1.2 meters. The subjects were not previously acquainted, and maintaining social distance may have led them to have only one person stand at the board. However, throughout the process, the dyad partners consulted each other strategically when it provided benefits, such as checking the data, or when they did not know which path to take. As Gattis (2002) describes, information seeking behavior adjusts to social behavior.

6. Conclusions

Collaborative information visualization behavior is a social process as well as a cognitive process. The process involves not just seeking and interpreting visualizations, but it also involves juggling complex activities including coordination, articulation, and social goals. Supporting coordination in particular remains a challenge to designers of CIVEs. The coordination mechanisms that we found groups used in our study, to establish roles, can be used to inform CIVE design. Support for roles can be a design feature to aid coordination.

Our results show that with a transparent system, users will wind up with a higher net productivity (taking the number, complexity and meaningfulness of all findings into account), but will achieve this gain with quite a bit of noise (i.e. non-meaningful findings). Unfortunately groups are no better than individuals in sifting through and eliminating trivial findings. We suggest, however, that with a more transparent system,

there is less coordination overhead for groups. Stages 1, 2, and 3 of the model, the planning stages, can help system designers understand where coordination problems lie in using a CIVE. Where groups seem to have the most advantage over individuals is in the validation stages: stage 4, identifying the solution and validating it, and stage 5, in the overall problem validation.

How can a model of the collaborative information visualization process be applied? First, the model can be used by usability engineers to understand where usage problems in a system lie. For example, if people take an inordinate amount of time in mapping the variables to the program (stage 2), then the variables in the system may be poorly represented. Second, the model can be used to compare systems. For example, if users take longer with one system in finding the correct visualization (stage 3), then this could indicate that visualizations are difficult to find or understand in the system. Last, the model can aid developers, to insure that they spend enough care in designing for each stage of the process. For example, being aware that people must re-interpret the data from visual form in the system into text form on paper (stage 5) can inform designers on how to ease this translation process.

We conclude with design recommendations that emerged from our study. Information visualization systems should be designed in such a way as to support the different stages in the information visualization process. We noted some system characteristics that seem to provide such support.

System transparency to support stages 2 through 4: The more transparent system (InfoZoom) made it easier for subjects to select variables (since nearly all of them were directly visible at the interface), and to loop through stages 2 – 4 (since they could do this

incrementally while Spotfire subjects had to plan the loops in advance). We noted similar effects in collaborative settings, which contribute to the faster task solution times with InfoZoom. As the choice of representation is critical to both free discovery and focused question tasks, transparency can aid users in selecting the representation.

Summarization to support stage 5: In this stage, subjects “retranslate” the visualization into natural language. We assume that system support in this retranslation will help analysts discover errors in the selected visualization more easily. InfoZoom features a status line that summarizes the selections made in zooming operations, which seems to be a first rudimentary step in this direction. Also, because we observed that validation is a collective act, the provision of rich media channels at this stage for remote work can be beneficial to support interaction.

Retracing steps to support stages 4 and 5: We discovered that it is valuable to retrace one’s steps through the system to validate. System support to provide a clear history of the information-seeking path can support not only validation but also can promote variable exploration. If subjects can view histories, then they can discover at what point their reasoning was faulty.

Finally, the placement of controls and size of the interface affect collaborative behavior. Subjects who were standing on the left-hand side in the collocated condition had much greater control over the visualization than subjects on the right-hand side, due to the fact that important control devices of the visualization were located on the left side. It remains to be seen whether the equal distribution of interface controls would lead roles to become more equal in face-to-face collaborative visualization. The size of the interface in collocated work can affect the degree of engagement. An interface must be of

sufficient scale so as to support multi-users comfortably. Wall projections can increase the scale but the tradeoff is losing direct interaction with the interface.

Much remains to be done in future work. Our experiment enabled us to abstractly model the environment, but it would be valuable to perform *in situ* studies to understand the organizational and social constraints that affect CIVE usage. Our model can also be applied with other systems to strengthen its validity. Though we have examined two common collaborative settings, our results may have been influenced by our choice of communication technologies. Though our quantitative results did not differ significantly, we did find important group process differences that we would like to further examine.

Acknowledgements

This research was supported by the National Science Foundation under grant no. 0093496, by the Center for Research on Information Technology and Organizations (CRITO), and by instrument grants from Smart Technologies Inc, humanIT AG and Spotfire Inc. We would like to thank Keri Carpenter for all her valuable work, and also Victor González, Richard Hyunh, Jeffrey Cheng, Cristina Gena and David Lim for helping with the experiments and their evaluation.

References

Ahlberg, C. & Wistrand, E. (1995). IVEE: An Information Visualization and Exploration Environment, *IEEE Symposium on Information Visualization*, 66-73.

Anupam, V. & Bajaj, C. L. 1993. Collaborative Multimedia Scientific Design in SHASTRA. *Proceedings of the first ACM international conference on Multimedia*, 447--456.

Anderson, J. R. (1980) *Cognitive Psychology and its Implications*. San Francisco, CA: W. H. Freeman and Co.

Berger, C. R. (1997). *Planning Strategic Interaction: Attaining Goals through Communicative Action*. Mahwah, NJ: Lawrence Erlbaum.

Blockeel, H. & Moyle, S. (2002). Centralized Model Evaluation for Collaborative Data Mining. *Conference on Data Mining and Warehouses (SiKDD 2002)*, Ljubljana, Slovenia, <http://www-ai.ijs.si/DunjaMladenec/SiKDD02/papers/BlockeelSep02.pdf>

Börner, K. (2001): A Collaborative Memory Palace for Digital Library Search Results: Usability Evaluation and Interface Design. In M. J. Smith, G. Salvendy, D. Harris & R. J. Koubek (Eds.), *Proceedings of the 2001 International Conference on Human-Computer Interaction*, London: Lawrence Erlbaum, 1, 160-1164.

Borgman, C. L. (1987). Individual differences in the use of information retrieval systems: Some issues and some data. *10th International ACM Special Interest Group on Information Retrieval Conference*, New Orleans, 61-71.

Bourne, P., Gribskov, M., Johnson, G., Moreland, J., Wavra, S., & Weissig, H. (1998). A prototype molecular interactive collaborative environment (MICE). *Proceedings of the Pacific Symposium on Biocomputing*, pp. 118-129.

Bowers, J., Pycock, J., & O'Brien, J. (1996). Talk and embodiment in collaborative virtual environments. *Proceedings of the conference on Human factors in computing systems (CHI'96)*, 58-65.

Bradner, E. & Mark, G. (2001). Social presence with video and application sharing. *Proceedings of ACM GROUP'01*, 154-161.

Brewer, I., MacEachren, A. M., Abdo, H., Gundrum, J. & Otto, G. (2000): Collaborative Geographic Visualization: Enabling Shared Understanding of Environmental Processes, *InfoVis 2000: IEEE Symposium on Information Visualization*, Salt Lake City, UT, 137-141.

Castells, M. (1996). *The Rise of the Network Society*. Blackwell: Oxford.

Chen, H., Nunamer, J., Orwig, R., & Titkova, O. (1998). Information visualization for collaborative computing. *IEEE Computer*, 75-82.

Chuah, M. C. & Roth, S. F. (2003). Visualizing common ground. *Proceedings of IEEE International Conference on Information Visualization (IV'03)*, 365-372.

Chi, E., Pirolli, P. & Pitkow, J. (2000). The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site. *Proceedings of the conference on Human factors in computing systems (CHI 2000)*. 161-168.

Churchill, E., Nelson, L., Denoue, L. & Girgensohn, A. (2003). The Plasma Poster Network: Posting Multimedia Content in Public Places. *Proceedings of INTERACT'03*.

Dix, A. J., Finlay, J. E., Abowd, G. D. & Beale, R. (2004). *Human-Computer Interaction*. Third Edition. Harlow, England: Prentice Hall.

Donahue, J. & Widom, J. (1986). Whiteboards: A Graphical Database Tool. *ACM Transactions on Office Information Systems*, 4(1), 24-41.

Edelson, D., Pea, R., & Gomez, L. (1996). Constructivism in the Collaboratory, In B. G. Wilson (Ed.): *Constructivist Learning Environments: Case Studies in Instructional Design*. Englewood Cliffs, NJ: Educational Technology Publications.

Elrod, S., Bruce, R., Gold, D., Goldberg, Halasz, F., Janssen, W., Lee, D., McCall, K., Pedersen, D., Pier, K., Tang, J., & Welch, B. (1992). LiveBoard: A large, interactive

display supporting group meetings, presentations and remote collaboration. *Proceedings of the conference on Human factors in computing systems (CHI '92)*, 599-607.

Gattis, L. (2002). Planning and Information Foraging Theories and Their Value to the Novice Technical Communicator. *ACM Journal of Computer Documentation*, 26(4), 168-175.

Gerald-Yamasaki, M. J. (1993). Cooperative visualization of computational fluid dynamics. *Computer Graphics Forum*, vol. 12(3), EG 93 Conference Issue, 497-508.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.

González, V. & Kobsa, A. (2003): Benefits of Information Visualization Systems for Administrative Data Analysts. *Proceedings of the Sixth International Conference on Information Visualisation (IV03)*, London, U.K. IEEE Press, Los Alamitos, CA, 331-336.
<http://www.ics.uci.edu/~kobsa/papers/2003-IV-kobsa.pdf>

Grave, M. (1994). Shared Data Spaces for Distributed Computing and Parallelism in Scientific Visualization Systems. *CWI Quarterly*, 7(2), 175-185,

Greenberg, S. (1990). Sharing views and interactions with single-user applications. *Proceedings of the ACM/IEEE Conference on Office Information systems*, 227-237.

Hall, E. T. (1966). *The Hidden Dimension*. Doubleday.

Halonen, D., Horton, M., Kass, R., & Scott, P. (1990). Shared hardware: a novel technology for computer support of face to face meetings. *Proceedings of the Conference on Office Information Systems 1990*, 163-168.

Isaacs, E. A. & Tang, J. C. (1993). What video can and can't do for collaboration: a case study. *Proceedings of ACM Multimedia, 1993*, New York: ACM Press, 199-205.

Jerding, D., Stasko, J., & Ball, T. (1997). Visualizing interactions in program executions. *Proceedings of the 19th International Conference on Software Engineering*, 360-370.

Johansen, R. (1988). *Groupware: Computer Support for Business Teams*. New York, NY, Free Press.

Johnson, J. & Johnson, P. (1993). Explanation facilities and interactive systems. *Proceedings of Intelligent User Interfaces '93*, 159-166.

Kobsa, A. (2001). An empirical comparison of three commercial information visualization systems, *IEEE Symposium on Information Visualization*, San Diego, CA, 123-130. <http://www.ics.uci.edu/~kobsa/papers/2001-INFOVIS-kobsa.pdf>.

Koenemann, J. & Belkin, N. (1996). A case for interaction: A study of information retrieval behavior and effectiveness. *Proceedings of CHI'96*, 205-212.

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 Conference*, San Diego, CA, 1999 <http://evlweb.eecs.uic.edu/aej/papers/astc99.pdf>.

Li, D. & Li, R. (2002). Transparent sharing and interoperation of heterogeneous single-user applications. *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 246-255.

Lorge, I. & Solomon, H. (1955). Two models of group behavior in the solution of Eureka-type problems. *Psychometrika*. 20, 139-148.

Mark, G., Carpenter, K., & Kobsa, A. (2003). A Model of Synchronous Collaborative Information Visualization. *Proceedings of IEEE Seventh International*

Conference on Information Visualization (IV'03), IEEE Press, 373-383.
<http://www.ics.uci.edu/~kobsa/papers/2003-IV-mark-kobsa.pdf>

Mark, G., Kobsa, A. and González, V. (2002). Do four eyes see better than two? Collaborative versus individual discovery in data visualization systems. *Proceedings of IEEE Sixth International Conference on Information Visualization (IV'02)*, IEEE Press, 249-255. <http://www.ics.uci.edu/~kobsa/papers/2002-IV-kobsa.pdf>

Mark, G. (2002). Conventions and commitments in distributed CSCW groups. *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, 11, (3-4), 349-387.

Muramatsu, J. & Pratt, W (2001). Transparent search queries: Investigating users' mental models of search engines. *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 217-224.

Mynatt, E., Igarashi, T., Edwards, W. & LaMarca, A. (1999). Flatland: New Dimensions in Office Whiteboards. *Proceedings of the 1999 International Conference on Human-Computer Interaction*, 346-353.

Newell, A. (1980), Reasoning, problem-solving, and decision processes: The problem space as a fundamental category, in R. Nickerson (ed.), *Attention and Performance VIII*, Hillsdale, NJ: Lawrence Erlbaum Associates, 693-718.

Norman, D. A. (1976). *Memory and Attention*. New York: Wiley.

Olson, G. M. & Olson, J. S. (2000). Distance matters, *Human-Computer Interaction*, 15, 139-179.

Pagendarm, H. G. & Walter, B. (1993). A prototype of a cooperative visualization workplace for the aerodynamicist. *Computer Graphics Forum*, 12(3), EG 93 Conference Issue, 485-496.

Palen, L. & Grudin, J. (1995). Why groupware succeeds: Discretion or mandate? *Proc. of the European Conference on Computer-Supported Cooperative Work (ECSCW'95)*, 263-278.

Pang, A. & Wittenbrink, C. M. (1997). Collaborative 3D Visualization with Cspray. *IEEE Computer Graphics and Applications*, 17, 32-41.

Park, K.S., Kapoor, A. & Leigh, J. (2000). Lessons learned from Employing Multiple Perspective in a Collaborative Virtual Environment for visualizing Scientific Data. *Proceedings of CVE' 2000*, 73-82.

Pedersen, E.R., McCall, K., Moran, T.P. & Halasz, F.G. (1993) Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings. *Proceedings of the 1993 International Conference on Human-Computer Interaction (CHI'93)*, 391-398.

Pirolli, P. & Card, S. (1999). Information foraging. *Psychological Review*, 106, 643-675.

Pirolli, P. & Card, S. (1995). Information foraging in information access environments. *Proceedings of the 1995 International Conference on Human-Computer Interaction (CHI '95)*, 51-58.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction*. New York: Wiley.

Raymond, E. S. (2003): *The Art of Unix Programming*. Boston, MA: Addison-Wesley.

Rogers, Y., Brignull, H. & Scaife, M. (2002). Designing dynamic interactive visualizations to support collaboration and cognition. *Proceedings of IEEE Sixth International Conference on Information Visualization (IV'02)*, IEEE Press, 39-48.

Russell, D., Trimble, J. & Wales, R. (2002). Two paths from the same place: Task driven and human-centered evolution of a group information surface. *Make IT Easy Conference*, June 2002.

Sawant, N., Scharver, C., Leigh, J., Johnson, A., Reinhart, G., Creel, E., Batchu, S., Bailey, S., & Grossman, R. (2000). The Tele-Immersive Data Explorer: A Distributed Architecture for Collaborative Interactive Visualization of Large Data-sets, *4th International Immersive Projection Technology Workshop*, Ames, Iowa, 2000 http://evlweb.eecs.uic.edu/cavern/TIDE/tide_ipt2000.pdf

Schönhage, B. (2000). *DIVA: Architectural Perspectives for Information Visualization*. Dissertation, Vrije Universiteit, Amsterdam, Netherlands.

Schwartz, S. (1971). Modes of representation and problem solving: Well evolved is half solved. *Journal of Experimental Psychology*, 91, 347-350.

Snowden, D. & Jaa-Aro, K. (1997). A subjective virtual environment for collaborative information visualization. *Virtual Reality Universe*. http://www.cyberedge.com/vru_papers/snowden.htm

Spenske, M. & Beilken, C. (1999). Discovery Challenge: Visual, Interactive Data Mining with InfoZoom – the Financial Data Set, Workshop Notes on Discovery Challenge, *3rd European Conference on Principles and Practice of Knowledge Discovery in Databases, PKDD'99*, 33-38. <http://fit.gmd.de/~cici/InfoZoom/DiscoveryChallenge/Financial.ps>

Spenske, M., Beilken, C. & Berlage, T. (1996). The Interactive Table for Product Comparison and Selection, *UIST 96 Ninth Annual Symposium on User Interface Software and Technology*, Seattle, pp. 41-50. <http://fit.gmd.de/~cici/Focus/Paper/uist96.htm>

Sproull, L. & Kiesler, S. (1991). *Connections: New Ways of Working in the Networked Organization*. Cambridge, MA: MIT Press.

Stefik, M., Foster, G., Bobrow, D.G., Kahn, K., Lanning, S. & Suchman, L. (1987). Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. *Communications of the ACM*, 30 (1), 32-47.

Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: Sage Publications.

Tang, J., Isaacs, E. & Rua, M. (1994). Supporting Distributed Groups with a Montage of Lightweight Interactions. *Proceedings of CSCW '94*. p. 23-34.

Tang, J. (1990). VideoDraw: A video interface for collaborative Drawing. *Proceedings of CHI'90*, 313-320.

Tatar, D. Foster & Bobrow, D. (1991). Design for conversation: lessons from Cognoter. *International Journal of Man-Machine Studies*, 34. 195-209.

Tromp, J., Bullock, A. Steed, Sadagic, A., Slater, M., & Frécon, E. (1998). Small Group Behavior Experiments in the COVEN project. *IEEE Computer Graphics and Applications*, 53-63.

Ulich, E., Rauterberg, M., Moll, T., Greutmann, T. and Strohm, O. (1991): Task Orientation and User-Oriented Dialog Design. *International Journal of Human-Computer Interaction* 3(2), 117-144.

Wood, J. (1998). *Collaborative Visualization*. Ph.D. Thesis, The University of Leeds, Leeds, England.

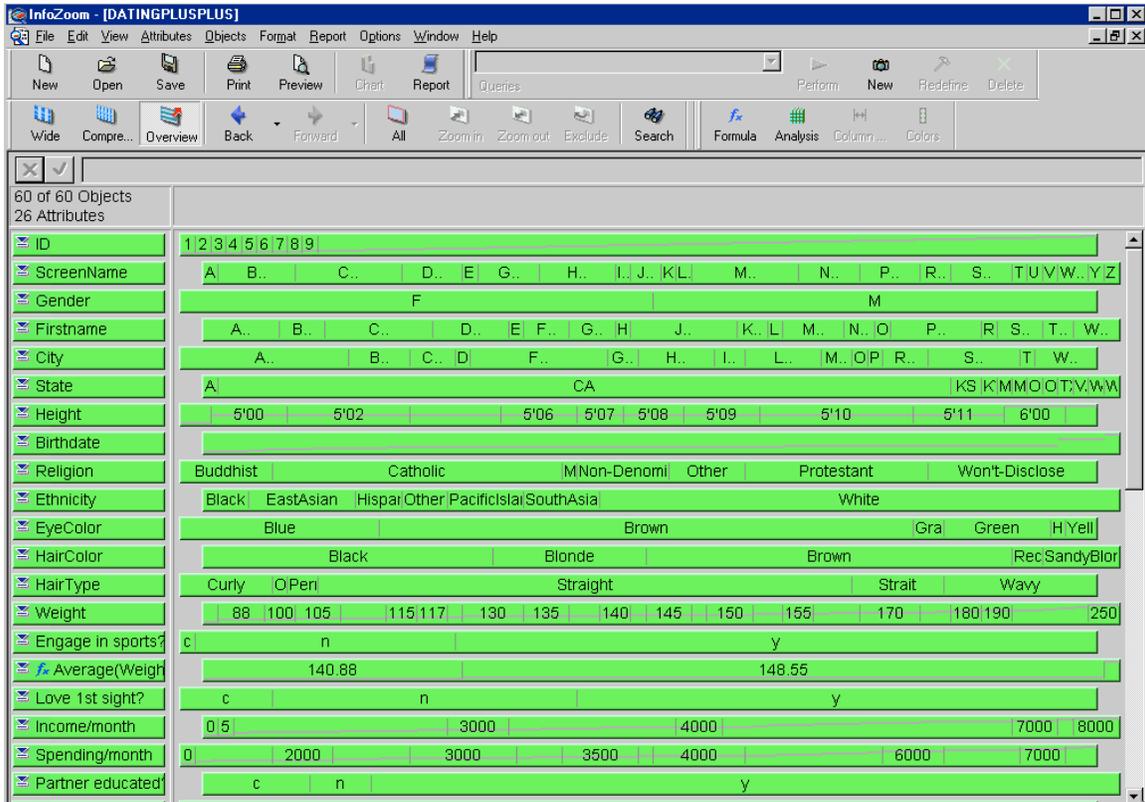


Figure 1

Weight		
Engage in sports?	n	y
Average(Weight)	140.88	148.55

Figure 2

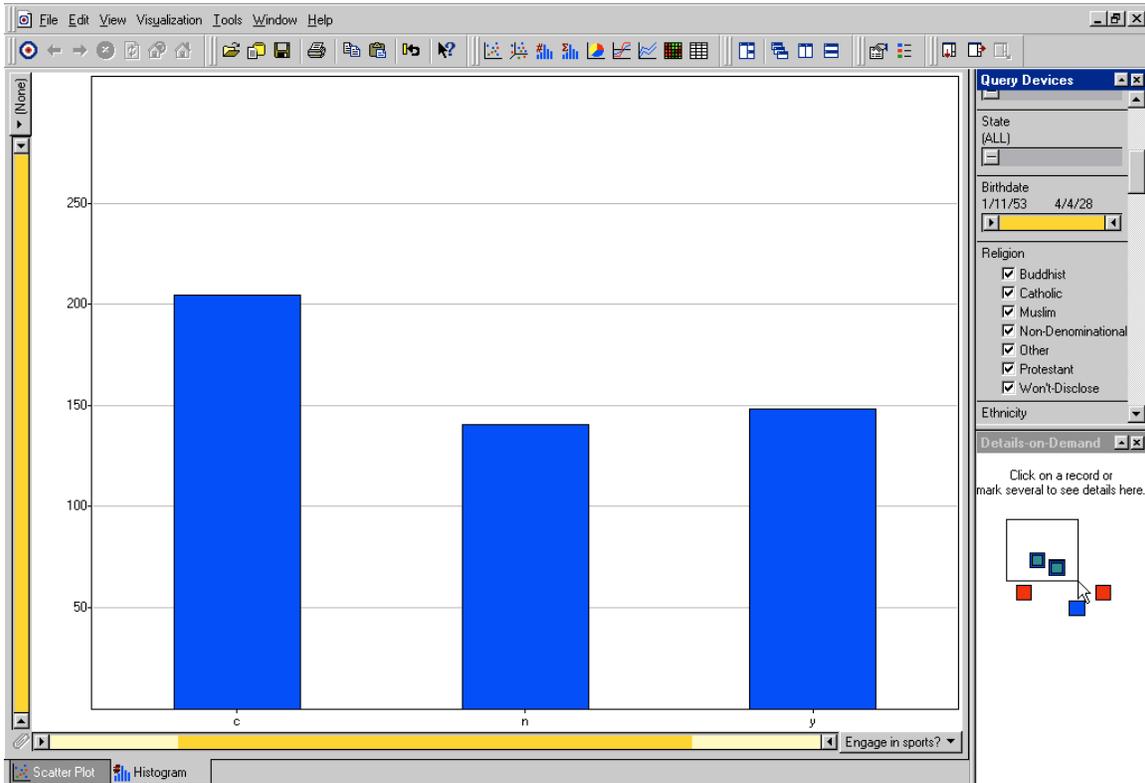


Figure 3

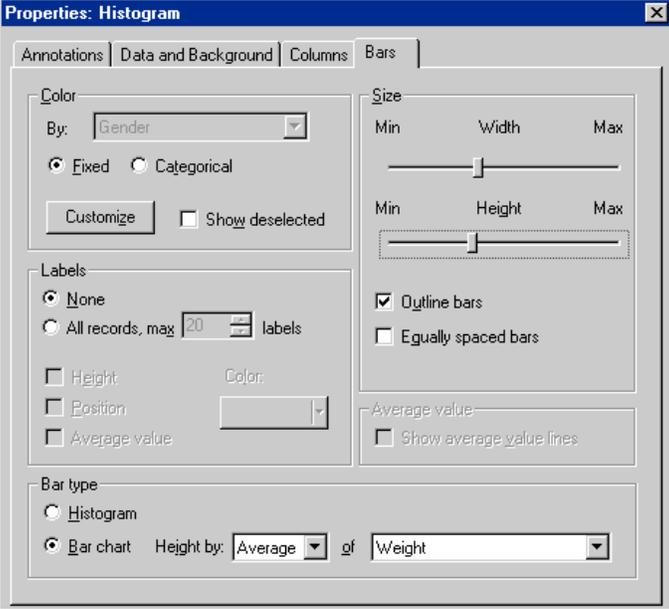


Figure 4

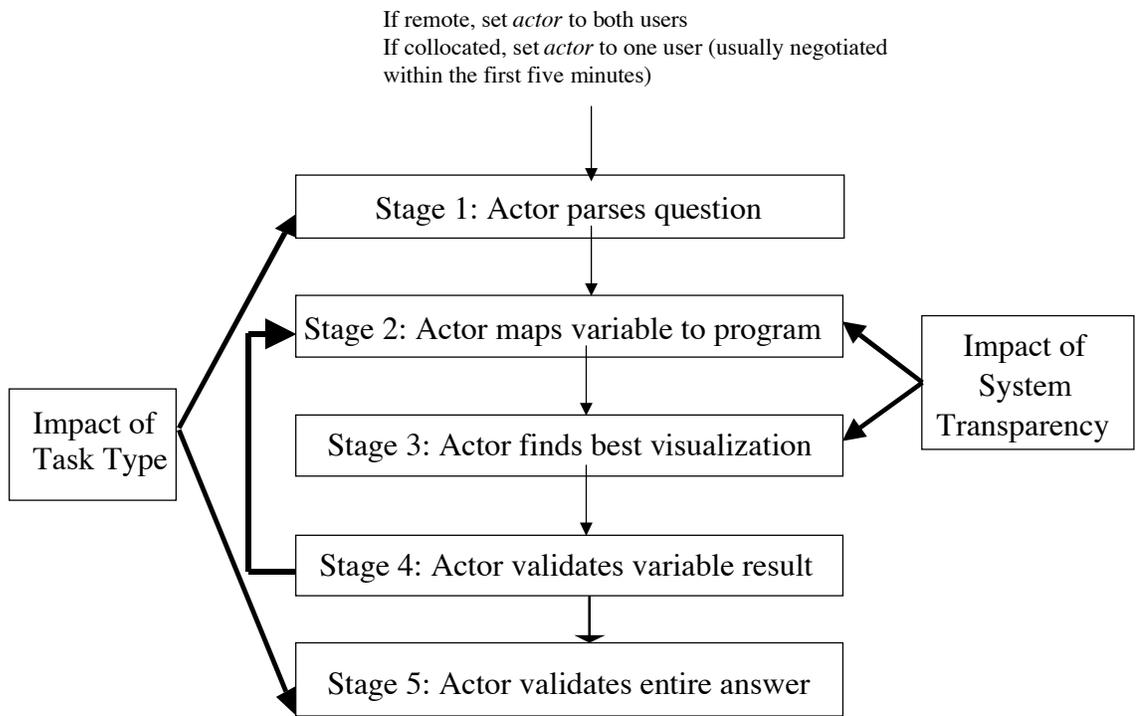


Figure 5

FIGURE CAPTIONS

Fig. 1: InfoZoom's overview mode

(figure 1: two-column width)

Figure 2: InfoZoom's compressed view

(figure 2: two-column width)

Fig. 3: A bar chart visualization in Spotfire

(figure 3: two-column width)

Fig. 4: Spotfire's "properties" menu used to set up the visualization in Fig. 3

(figure 4: two-column width)

Figure 5. A model of the collaborative visualization process.

(figure 5: two-column width)