UNIVERSITY OF CALGARY

Collaborative Information Visualization in Co-located Settings

by

Petra Isenberg

PhD PROPOSAL

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA
November, 2007

Abstract

It is common for small groups of people to gather around visual displays of information to discuss or interpret the information to form decisions. Groups can share the task load of exploring large and complex datasets and can share various interpretations of a dataset when working together. However, tools to support synchronous collaboration between several co-located people in their data analysis are still relatively scarce. Traditionally, information visualization tools have been designed from a single-user perspective. Research on collaborative data analysis has just recently received increased research attention and primarily distributed data analysis tools have been developed.

The design of digital systems for co-located synchronous collaboration around information visualizations poses challenges that have not been considered in single-user information visualization systems. In information visualization, it is not yet understood (1) how people collaboratively work with visual representations of data and which methods they use to solve information analysis tasks as a team, and (2) how work on other co-located collaborative activities (e. g. collaborative photo sorting, document editing, games) applies to the specific problem of collaborative data analysis. There are also only few examples of co-located collaborative data analysis software (e. g. VERNIER et al. (2002)) and few descriptions of collaborative data analysis practices in real-world environments from which to draw advice on how to design collaborative information visualization systems.

The research goal of this dissertation is to inform the design of information visualization tools to support co-located collaborative data analysis and to further our understanding of how people work together over information displays. In this proposal, I describe more specifically the research challenges I intend to address during my dissertation and present how my past work relates to these.

Contents

1 Introduction

Humans have been creating visualizations for thousands of years. Examples range from early maps, scientific drawings, or data plots to the interactive digital representations of large information spaces that we create today. We collect more information through our visual sense than through all other senses combined (WARE, 2000) and it is not surprising that many disciplines rely on visual representations of data to make a discovery or to communicate a discovery in a dataset.

By additionally providing data representations in a manipulable medium, varying parameter spaces can be interactively explored. The research field of information visualization is concerned with the creation of such interactive visual representations of abstract, nonphysically-based data. The goal is to help in making discoveries in a dataset, forming decisions based on insight gained from analyzing a dataset, explaining a phenomenon based on an insight, or to help in predicting future trends (CARD et al., 1999; WARE, 2000). This stands in contrast to the definition of scientific visualization which has essentially the same goal but largely operates on physically-based data, including techniques such as volume rendering or flow visualization.

So far, research in information visualization has largely focused on supporting a single person in doing data analysis. However, the process of analyzing and interpreting information and making decisions based on an analysis is often collaborative in nature, in particular, when important decisions are based on large and complex data sets (CHUAH and ROTH, 2003). The notion of data analysis as a social process raises interesting research issues for supporting collaboration around interactive digital information visualizations. In this dissertation proposal, I present a research plan to investigate several of the challenges of providing computational support for collaborative data analysis using information visualization.

1.1 Research Motivation

Humans have considerable experience collaborating in shared work environments and collaborative data analysis also often occurs in co-located settings. Decisions and conclusions are rarely based on the analysis by a single person (THOMAS and COOK, 2005). Team members can offer different perspectives and expertise that together can improve the quality of decisions or solutions. For example, imagine a team of medical practitioners meeting to discuss a patient's medical record (Figure 1a), a team of scientists coming together to argue about the results of their latest experiments (Figure 1b), or a team of business analysts negotiating next year's budget based on the analysis of a financial data set (Figure 1c).

(a) Two nurses discussing a medical visualization.

(b) A group of astronomy students analyzing visual representations from their latest experiment.

(c) A team meeting to discuss a business plan.

Figure 1 – Examples of co-located collaborative data analysis over information displays.

Finding ways to augment this type of collaborative data analysis with the power of digital information visualizations may lead to more effective decision making. Most techniques so far have been designed to support this effective decision making for a single analyst by providing new techniques to conquer problems such as those of displaying increasingly large and multi-dimensional data sets (e. g. JERDING and STASKO (1995); MUNZNER et al. (2003)), finding appropriate visual support for relational information like hierarchies, clusters, temporal trends, outliers (e. g. ROBERTSON et al. (1991); VAN WIJK and VAN SELOW (1999)), or providing appropriate interaction techniques to explore complex datasets (e. g. CARPENDALE and MONTAGNESE (2001); VAN WIJK and NUIJ (2003)).

Often neglected in this research is the notion that data analysis is also commonly a social experience. With large data sets, the task load of exploring the data could be shared among several individuals on a team (THOMAS and COOK, 2005). Datasets on which decision and discoveries are based may also be susceptible to a variety of interpretations, in which case experts may discuss and negotiate their interpretations of the data.

Motivated by these benefits of collaborative data analysis, this dissertation investigates the challenges of supporting data analysis for *co-located, synchronous* work environments.

1.2 Research Scope

The topic of this dissertation lies at the intersection of two research fields as seen in Figure 2: Information Visualization (InfoVis) and Computer Supported Cooperative Work (CSCW).

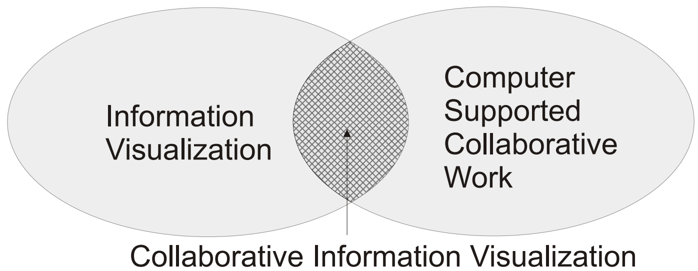


Figure 2 – Research Context.

In the field of information visualization, researchers have been working towards developing new visual representations, presentation and interaction techniques to amplify human cognition for different types of datasets, tasks, and analysis scenarios (CARD et al., 1999; CHEN, 2006; SPENCE, 2007). Research from the field of information visualization informs the topics of this dissertation by providing information on how individuals work with and perceive visual data representations, how they perform data analysis, and how to design interactive information visualization systems to support these work processes.

The field of CSCW (DIX et al., 1998, Ch. 13) is concerned with the challenges of designing software for multiple users to work as a group and how to understand the effect of deployed software on their work processes. Within CSCW, in particular the work on co-located groupware has a high applicability to the research problems of how to support co-located collaborative data analysis. Techniques that describe mechanisms to support coordination of activities in the workspace, awareness of group member's activities, access to and transfer of items in the workspace, for example, can be applied to the design of collaborative systems for co-located data analysis (ISENBERG and CARPENDALE, 2007).

Since an exhaustive investigation into all aspects of co-located collaborative information visualization is beyond the scope of this thesis and because there is little research in information visualization on the support of co-located data analysis, this research starts with the fundamentals in the research area based on specific data sets and tasks.

In particular, I conduct one of the first investigations into the difference between individual and collaborative data analysis processes, develop the initial set of guidelines for the design of co-located collaborative information visualization, and investigate the collaboration practices around information visualization in a scientific research laboratory in which a digital system for co-located data analysis will be deployed.

2 Foundations

In this section, I discuss research areas relevant to my dissertation. I start by reviewing research in information visualization on analysis processes and guidelines for design and introduce research within CSCW on co-located synchronous collaboration. Then, I describe the small body of research on collaborative data analysis and visualization systems in the synchronous and asynchronous distributed domains.

2.1 Information Visualization

To understand collaborative analysis processes, we need to first look at how individuals work with information visualizations. Several researchers have outlined frameworks that describe the use of information visualizations to solve problems. These frameworks share the common characteristic of modeling a person's involvement in the visualization process as an iterative sequence of components; however, each model is unique in terms of its focus, and how it abstracts the process. Several researchers describe analysis practices through operations in digital information visualization systems (e. g. (SHNEIDERMAN, 1996; CHI and RIEDL, 1998; JANKUN-KELLY et al., 2007)), while others focus on descriptions of cognitive analysis activities or tasks, (RUSSELL et al., 1993; CARD et al., 1999; AMAR and STASKO, 2005; THOMAS and COOK, 2005). For the focus of this dissertation, the last group of research on cognitive analysis and tasks is highly relevant to help inform an understanding of the differences and similarities to collaborative analysis practices.

Increasing our understanding of how data analysis is performed, provides insight into the design interactive visualizations that support this process. Several heuristics exist from which one can draw advice on the design of interactive visual representations, however the heuristics often differ in focus (ZUK et al., 2006). For example, the focus may be on a certain data domain, e. g. for ambient displays (MANKOFF et al., 2003), interaction (Baldonado et al., 2000), task support (Shneiderman, 1996), or based on perception and cognition (Ware, 2000; Amar and Stasko, 2005). Heuristics for the design of information visualizations can be used to inform the development of guidelines for collaborative information visualization systems based on the chosen tasks and datasets.

2.2 Co-located Synchronous Collaboration

Humans have considerable experience working together with others in a shared space at the same time. In the area of CSCW, several approaches have been described that have been designed to support synchronous co-located collaboration with technology. This technology can be in the form of large single-display technology like interactive wall (e. g. (Guimbretière, 2002, Ch. 4)) or tabletop displays (e. g. Wellner (1993)), or in the form of integrated mobile and wireless devices (e. g. Johnson et al. (2002)). To narrow the focus of this thesis, I will concentrate on collaboration around single-display technology (Stewart et al., 1999). Research in this area has, for example, described mechanisms to support coordination of activities in the workspace (e. g. Nacenta et al. (2007); Ringel Morris et al. (2004); Scott et al. (2004)), awareness of group member's activities (e. g. (Tang et al., 2006)), access to and transfer of items in the workspace (e. g. (Kruger et al., 2004)). As part of my proposed research, I have described several aspects of how research within CSCW, and in particular co-located collaboration, can be applied to the design of co-located collaborative information visualization systems (Isenberg and Carpendale, 2007). Details can be found in the Section 4 and in the Appendix.

2.3 Collaborative Information Visualization

Research on the process of *collaborative* data analysis using information visualizations is relatively scarce. Mark et al. (2002, 2003) conducted a user study in which they observed pairs working in co-located and distributed settings with two different visualization systems designed for single users. Their findings suggest that the benefit of collaborative vs. individual problem solving was heavily dependent on the visualization system used but that, in general, groups were better at locating errors (Mark et al., 2002). In their second paper, Mark et al. (2003) introduce a model for the collaborative problem-solving process. The model consists of an iterative sequence of five stages: parsing a question, mapping variables to the program, finding the correct visualization, and two validation stages. The second analysis of this study cautioned that for co-located collaboration, the placement of system controls affected the roles that collaborators were taking on during their shared work. This finding was highly influenced by their study setup as groups had to negotiate their interactions through a single input device in systems design for individual use on a desktop screen.

2.3.1 Distributed Visualization Systems

Most of the research into collaboration around information visualization has focused on distributed data analysis. CoMotion is a collaborative environment for creating information analysis and decision-support applications (MayaViz, 2007). The application provides shared views of the data on which all users can *synchronously* interact. The Command Post of the Future is a visualization tool built on this architecture in which distributed team members can share visualizations of natural emergencies and combat situations (Chuah and Roth, 2003). DecisionSite Posters (Spotfire, Inc., 2007), is a web-based system for asynchronous collaborative work around information visualization. It allows users to publish visualization results, descriptions, and data to distributed collaborators in interactive web-based reports. Many Eyes (Viégas et al., 2007), and Swivel[1] are two systems that are targeted at an internet-scale audience and both collaborative sharing and exploration of data by letting users upload and visualize data as well as comment on created visualizations.

While my research is not directly concerned with distributed collaborative visualization, this work will help to inform my research where relationships can be drawn between information visualization in distributed and co-located settings.

2.3.2 Co-located Visualization Systems

The responsive workbench was one of the first visualization systems for co-located collaboration around a large horizontal surface (Wesche et al., 1997). The responsive workbench is a virtual reality environment in which the displayed 3D scene is looked at through shuttered glasses. Several scientific visualization applications were developed for this platform including fluid dynamics or flow visualizations and situational awareness applications.

1 http://www.swivel.com/

There are fewer systems which have attempted to provide an environment specifically for co-located collaboration around information visualizations. With a focus on interaction, radial tree layouts have been studied for collaborative circular tabletop systems (VERNIER et al., 2002). In their system Vernier et al. provide two different fisheye mechanisms to support different types of user activity around the tabletop. Interactive focus+context techniques for collaborative systems have further been applied in the DTLens system by FORLINES and SHEN (2005). Initial work done as part of this dissertation involved first providing a set of guidelines for the development of collaborative information visualization systems and then creating a system design based on these guidelines (ISENBERG and CARPENDALE, 2007).

2.4 Summary

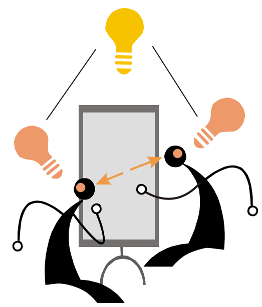
Research in information visualization draws from the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art (MUNZNER, 2000). The synthesis of relevant ideas from these fields is critical for the design and evaluation of information visualization in general and it is only sensible to think that research concerned with collaborative work also adds valuable information to our understanding of requirements for collaborative information visualization systems.

During my dissertation I will work towards extending our understanding of requirements for the design of co-located collaborative information visualization systems by integrating appropriate research from several fields as discussed above. Further information on progress in this direction can be found in Section 4.

3 Research Overview

The design of collaborative systems poses challenges that have not previously been considered in single-user information visualization. In a group setting the use of co-located collaborative technology needs to support a process of social interaction around the data, ideally, helping the group to arrive at a *common* understanding of the data through a process of collaborative interpretation, analysis, discussion, and interaction. That is, using these tools, groups can gain more than the simple combination of two persons' individual insight from the data. The different goals of individual and collaborative information visualization are summarized in Figure 3. In Figure 3a a single person arrives at an insight or discovery through a process of looking at and possibly interacting with an information display, forming a mental model and interpreting the data display, and ideally gaining an insight (Spence, 2007). In Figure 3b two people join in a collaborative analysis. They both come to individual insight by looking at and interpreting the dataset but through the social interaction (discussion, negotiation, interaction) they both can reach a common understanding of the dataset which may lead to more effective decision making.

(a) A single analyst looks at the digital visualizations, engages in cognitive processing about the visual data display, and, ideally, arrives at an insight or discovery about the data set.



(b) Two analysts join in a collaborative analysis scenario. The goal is that both will come to a common understanding of the data (group cognition) through visualization use.

Figure 3 – Goals of single-user and collaborative information visualization.

The challenge in designing information visualization for this type of synchronous and co-located collaborative work is that mechanisms need to be designed that support the ways people work *together* during an analysis. It is still relatively unexplored how to design these systems so that they support the generation of a common understanding through collaborative interaction with and analysis of information visualizations. This thesis does not attempt to describe all the ways that technology does or could impact upon synchronous co-located collaborative data analysis but instead proposes to shed light on a specific set of research problems. This includes a literature review of related CSCW research, an investigation into collaborative analysis practices, the implementation of a specific collaborative system, and a study of collaborative software with domain experts in a microbiology research lab (see Figure 4).

We need to consider how to apply previous research on collaborative technology to the problem of co-located collaboration around information visualization. In particular, we need to consider previous studies for the design of software for co-located col-

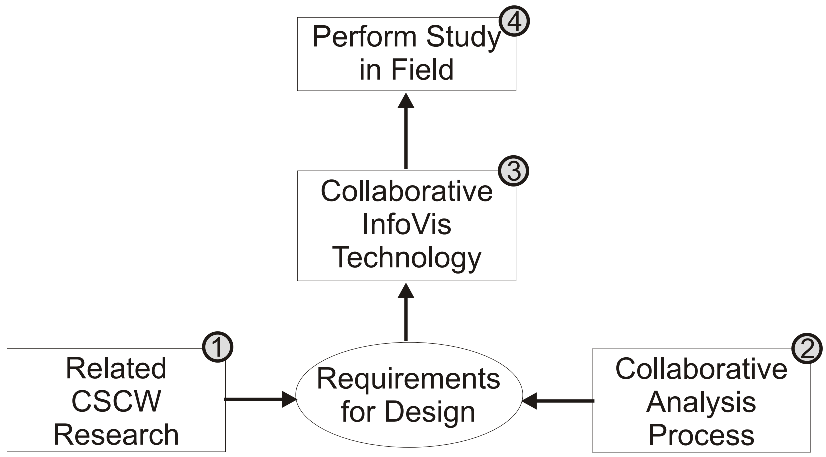


Figure 4 – Research Proposal.

laboration (e. g., GUTWIN and GREENBERG (1998); KRUGER et al. (2004); RINGEL MORRIS et al. (2004); SCOTT et al. (2004); TANG et al. (2006)). However, work around information visualizations such as discovery and analysis tasks, differs from other collaborative work scenarios like design projects, information organization (e. g. photo sorting), or document editing in several ways. For one, the outcome of an information analysis is not a product (design, organized photo collection, or an edited document) but is an understanding or insight. Other possible issues arise because information visualizations have both an interaction component and a data representation component. Both of these components may need rethinking and redesigning. How general guidelines and research on the design of systems for other co-located collaborative work apply to data analysis scenarios is an open problem. Part of my research plan is to do an in-depth literature survey to see how and where research on co-located and distributed synchronous collaboration applies to synchronous co-located data analysis scenarios (① in Figure 4). This analysis will lead to an initial set of design guidelines for these systems (③ in Figure 4). Progress in this direction is summarized in Section 4.

One of the design challenges for co-located collaborative information visualization systems is that we do not yet have a clear understanding of the information analysis methods used during collaborative work. In order to design digital information visualization systems that can adequately support collaborative work, we need to investigate *how* people collaboratively analyze information. How are information visualizations used by teams? How *could* teams use information visualizations in their collaborative process? For example, how a single doctor analyzes biomedical information visualizations might differ from how a team of doctors analyzes the same data. While many researchers have explored the information analysis process (e. g., CARD et al. (1999); JANKUN-KELLY et al. (2007); SPENCE (1999)), little has emerged on the nature of these methods in a collaborative context (e. g., MARK et al. (2003)). I plan to investigate the differences between how individuals and small co-located teams (e. g., two to three individuals) make use of visual information during collaborative work, starting with an investigation into how analysis is conducted in a natural, non-digital setting (see ② in Figure 4 and Section 4).

The result of the first two phases of my proposed research can help to form a set of initial design requirements for digital co-located collaborative information visualization systems. Starting from these requirements (③ in Figure 4) I propose to design a system for co-located collaborative work around interactive information visualizations (③ in Figure 4). This system will be designed for an expert user group, a microbiology lab at the University of Calgary. Using this system I will perform a field study with the goal of increasing my understanding of the requirements of technology use in a real-world environment. I plan to first understand how and when collaborative data analysis is currently conducted in their research environment and then iteratively design software to support aspects of this work process. The information gained from such an evaluation will inform our understanding of the use of digital collaborative software for data analysis in an example work settings (④ in Figure 4).

4 Contributions Thus Far

To this point in my research, I have contributed:

- a set of derived design guidelines for co-located collaborative information visualization systems from an in-depth literature review, and
- the articulation of eight processes common to collaborative data analysis that need to be supported in co-located collaborative information visualization tools.

The initial set of design heuristics is derived from a literature review of three research areas: information visualization, distributed and co-located collaboration advice, and the studies that look directly at collaborative visualization. The intention is that these design guidelines will form a basis which will adjust and expand as research in collaboration around information visualizations continues. The set of guidelines is one of the first specifically tailored towards the design of digital information visualization systems for co-located synchronous data analysis and has been published in Transactions on Visualization and Computer Graphics (Proceedings of InfoVis) (ISENBERG and CARPENDALE, 2007).

These design guidelines were used[2] to develop a digital system to support different working styles around information visualizations on digital tabletop displays (ISENBERG and CARPENDALE, 2007). I developed a collaborative environment in which hierarchical data can be explored and compared through the use of several features that have been designed to facilitate collaborative work practices: multi-user input, shared and individual views on the hierarchical data visualization, flexible use of representations, and flexible workspace organization. One example analysis scenario using my system can be seen in Figure 5. In this case, I focused on a specific data type (hierarchical data sets) and task (hierarchical data comparison), and a clear next step would be to understand how existing single-user representation and visualization interaction techniques should be modified to fulfill the needs of analysis teams.

Figure 5 – Use of my collaborative tree comparison software.

One of the problems uncovered during the initial literature review and design of a collaborative software prototype was that we do not have a good understanding of how groups analyze information together in a shared workspace and how information visualization are read and interacted with during analysis. A knowledge of these processes is important to adequately support these analysis practices with software.

To inform our understanding of collaborative data analysis practices, I conducted[3] an observational study to understand the visual analysis process for small groups

2 Co-author on this work: Sheelagh Carpendale
3 Co-authors on this work: Anthony Tang and Sheelagh Carpendale

compared to individuals (NEUMANN et al., 2007). We decided to observe participants' natural working styles, unencumbered by any specific digital interface (see Figure 6). This setup allowed the observation of people's approach to group analysis of visual information including behaviours such as free arrangement of data, annotation practices, and different ways of working with individual information artefacts—behaviours that would not otherwise be observable given most digital information visualization tools. A key drawback of this approach is that we would not see how typical interactions in information visualization tools (such as selection, encoding, or presentation parameter manipulations) would be used; however, our specific interest was in uncovering the *general processes* involved in collaborative and individual visual analysis.

The analysis of our observations revealed eight processes common to how participants completed the tasks in our study. We have shown how these eight processes relate to other models of information analysis, and provided insights on differences and commonalities between them. Yet, while others have posited a general temporal flow of information analysis, our results suggest this temporal flow may simply reflect an assumption in the design of existing information visualization tools. Thus, we argue that designers should consider individuals' unique approaches toward analysis by supporting a more flexible temporal flow of activity.

Figure 6 – Users explore data collaboratively during an observational study.

5 Future Contributions

For the remainder of my PhD I will explore data analysis in the context of a biology research lab. I have started a collaboration with Mike Surette's Bacterial Pathogenesis Research Group at the University of Calgary that currently performs collaborative analysis of data derived from their own experiments. Within their environment I plan to conduct three main research activities:

1. investigation on the types of collaborative data analysis used within this research lab
2. development of a software prototype for co-located collaborative data analysis
3. study this prototype system with expert users from this research group

Investigation on Collaborative Analysis Activities:
My initial investigations indicate that several different types of synchronous co-located collaborative analysis practices are used: (a) *casual* co-located data analysis analysis, for example when one person casually asks another person to take a look at the results from his or her latest experiment; and (b) *formal*, synchronous, co-located data analysis activities, for example, when students, post-docs, and faculty join in a planned labmeeting to discuss the results of an experiment.

I intend to continue my studies and investigation in order to learn how this list can be extended and refined. I want to find out at which stages during the analysis of an experimental dataset co-located data analysis is conducted, for which types of questions, and in which form. This investigation will inform my understanding of where and how the different types of co-located data analysis can be supported with technology. While there are many methodologies I could use (such as different types of observations, interviews, focus groups, etc. (Denzin and Lincoln, 2005)) to approach this study, I am still considering which will be the most appropriate methodology. I will likely choose some combination of contextual inquiry, in-depth interviews, and questionnaires.

Prototype Development:
Initial discussions with biologists revealed that they currently have no software support for co-located collaborative data analysis. That is, with their current software it is not possible for several people to interact with the data displays. Their data analysis currently involves the use of several disconnected software tools. For example, first Excel is used to format the data, and then the result is used in a program for cluster analysis. The cluster result is then used as input for another program that creates a visual display of the results. From these results, static images are created for casual or formal discussions. During the formal discussions I observed, lab members frequently expressed the wish to discuss a representation of the data based on different parameters (e. g. a different cluster algorithm) or using a different data representation. This would require the presenter to retrace their analysis steps and create a new representation to include in their presentation and this is typically left for another meeting.

During a demonstration of my initial research prototype (see Section 4) several biologists stated that such a system could greatly enhance their data analysis practices. My prototype, however, since it was designed based on higher-level interaction and collaboration advice, is not specific enough in that it does not support the specific data and analysis tasks important to this research group.

The knowledge from directly studying collaborative data analysis practices as they actually take place in the biology research lab plus the knowledge from my previous research will inform the design of a digital system that helps to support a more interactive collaborative discussion of their data graphics. The software for this system will be designed in collaboration with Matthew Tobiasz, a MSc student in the iLab. We are currently in the process of understanding the typical data sets that the biologists analyze and which tasks are involved in drawing conclusions from the data. We intend to further involve our expert user group in a user-centered design approach. Insight gained from their involvement during the initial design exploration and problem definition will be used to direct the system development and will help to evaluate proposed solutions. The task of developing the analysis software will be divided between me and Matthew in a way that individual contributions are clearly separable. While we both are taking part in the participatory investigations, Matthew will focus on the development of the representation of the biologist's data and I will focus on the changes, adjustments, interactions, and widgets required to enable the observed collaborative practices.

Investigation into Software Use:
Once a first usable prototype has been developed, I will investigate how this digital system will be used and how it influences or changes the biologists' way of collaborating around data. In particular, I am interested in increasing my understanding of the eight analysis processes uncovered in my previous work (see Section 3) and exploring and evaluating whether and how our software supports these eight analysis processes. Insights will help to improve the software design and will shed light on how participants engage in the eight processes using interactive digital information visualizations.

6 Timeline

The following timeline (Figure 7) shows my past research and future plans for completing the remaining phase of my research. As indicated above, I have conducted an initial literature review in the summer of 2006 into research that can help to inform design of co-located collaborative information visualization systems. I expect this review to expand as my research continues. In Fall 2006 I implemented an initial research research prototype. The pen and paper based study of collaborative analysis process was designed, conducted, and analyzed during Winter and Spring 2007. In the Fall of 2007 I have begun initial explorations into the work processes at a biology research lab. These preliminary results will be used to design a formal study of the analysis processes used in their research environments and inform the implementation of a co-located collaborative software environment in the Winter and Spring of 2008. This software will in turn be analyzed with the domain experts during the Summer of 2008. Then, I plan to spend two semesters writing my thesis in order to defend by Spring 2009.

Summer 2006	Literature Review
Fall 2006	Implementation of Prototype
Winter/Spring 2007	User Study of Collaborative Analysis Process
Fall 2007	Initial Field Study for Technology Deployment
Winter/Spring 2008	Microbiology Analysis Study and Technology Development
Summer 2008	User Study in Biology Research Lab
Fall/Winter 2008	Thesis Writing
Spring 2009	Thesis Defense

Figure 7 – Proposed thesis timeline.

Appendix

- Accepted paper on literature review and prototype design: in *IEEE Transactions on Visualization and Computer Graphics 12(5)*.
- CHI 2008 submission on study of visual analysis process.

References

ROBERT A. AMAR and JOHN T. STASKO. Knowledge Precepts for Design and Evaluation of Information Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):432–442, 2005.

MICHELLE Q. WANG BALDONADO, ALLISON WOODRUFF, and ALLAN KUCHINSKY. Guidelines for Using Multiple Views in Information Visualization. In *Proceedings of AVI '00*, pages 110–119, New York, USA, 2000. ACM Press.

STUART CARD, JOCK D. MACKINLAY, and BEN SHNEIDERMAN, editors. *Readings In Information Visualization: Using Vision To Think*. Morgan Kauffman Publishers, Inc., San Francisco, USA, 1999.

M. SHEELAGH T. CARPENDALE and CATHERINE MONTAGNESE. A Framework for Unifying Presentation Space. In BILL SCHILIT, editor, *Proceedings of ACM Symposium on User Interface Software and Technology (UIST'01)*, pages 61–70, New York, NY, USA, 2001. ACM Press.

CHAOMEI CHEN. *Information Visualization: Beyond the Horizon*. Springer-Verlag, London, 2nd edition, 2006.

ED HUAI-HSIN CHI and JOHN T. RIEDL. An Operator Interaction Framework for Visualization Systems. In *Proc. of the Symposium on Information Visualization (InfoVis)*, pages 63–70, Los Alamitos, USA, 1998. IEEE Comp. Society.

MEI C. CHUAH and STEVEN F. ROTH. Visualizing Common Ground. In *Proceedings of the Conference on Information Visualization (IV)*, pages 365–372, Los Alamitos, USA, 2003. IEEE Computer Society.

NORMAN K. DENZIN and YVONNA S. LINCOLN, editors. *Handbook of Qualitative Research*. Sage Publications, Inc., Thousand Oaks, London, New Delhi, 3rd edition, 2005.

ALAN J. DIX, JANET E. FINLAY, GREGORY D. ABOWD, and RUSSELL BEALE. *Human-Computer Interaction*. Prentice Hall, London, New York, and others, 1998.

CLIFTON FORLINES and CHIA SHEN. DTLens: Multi-user Tabletop Spatial Data Exploration. In *Proceedings of the Symposium on User Interface Software and Technology (UIST)*, pages 119–122, New York, USA, 2005. ACM Press.

FRANÇOIS GUIMBRETIÈRE. *Fluid Interaction For High Resolution Wall-Size Displays*. Phd thesis, Stanford University, 2002.

CARL GUTWIN and SAUL GREENBERG. Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. In *Proceedings of Computer Supported Cooperative Work (CSCW)*, pages 207–216, New York, USA, 1998. ACM Press.

PETRA ISENBERG and SHEELAGH CARPENDALE. Interactive Tree Comparison for Co-located Collaborative Information Visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2007)*, 12(5):1232–1238, 2007.

T.J. JANKUN-KELLY, KWAN-LIU MA, and MICHAEL GERTZ. A Model and Framework for Visualization Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):357–369, 2007.

DEAN F. JERDING and JOHN T. STASKO. The Information Mural: A Technique for Displaying and Navigating Large Information Spaces. In NAHUM GERSHON and STEVE EICK, editors, *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '95)*, pages 43–50, Los Alamitos, CA, USA, 1995. IEEE Press.

BRAD JOHNSON, ARMANDO FOX, and TERRY WINOGRAD. The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms. *IEEE Pervasive Computing Magazine*, 1(2):67–74, 2002.

RUSSELL KRUGER, SHEELAGH CARPENDALE, STACEY D. SCOTT, and SAUL GREENBERG. Roles of Orientation in Tabletop Collaboration: Comprehension, Coordination and Communication. *Journal of Computer Supported Collaborative Work*, 13(5–6):501–537, 2004.

JENNIFER MANKOFF, ANIND K. DEY, GARY HSIEH, JULIE KIENTZ, SCOTT LEDERER, and MORGAN AMES. Heuristic Evaluation of Ambient Displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 169–176, New York, NY, USA, 2003. ACM Press.

GLORIA MARK, KERI CARPENTER, and ALFRED KOBSA. A Model of Synchronous Collaborative Information Visualization. In *Proceedings of the Conference on Information Visualization (IV)*, pages 373–381, Los Alamitos, USA, 2003. IEEE Computer Society.

GLORIA MARK, ALFRED KOBSA, and VICTOR GONZALEZ. Do Four Eyes See Better than Two? Collaborative versus Individual Discovery in Data Visualization Systems. In *Proceedings of the Conference on Information Visualization (IV)*, pages 249–255, Los Alamitos, USA, 2002. IEEE Computer Society.

MAYAVIZ. Comotion®. Website, 2007. http://www.mayaviz.com/ (accessed March 2007).

TAMARA MUNZNER. *Interactive Visualization of Large Graphs and Networks*. Phd thesis, Stanford University, California, USA, 2000.

TAMARA MUNZNER, FRANÇOIS GUIMBRETIÈRE, SERDAR TASIRAN, LI ZHANG, and YUNHONG ZHOU. TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility. *ACM Transactions on Graphics*, 22(3):453–462, 2003.

MIGUEL A. NACENTA, DAVID PINELLE, DANE STUCKEL, and CARL GUTWIN. The Effects of Interaction Technique on Coordination in Tabletop Groupware. In CHRISTOPHER G. HEALEY and EDWARD LANK, editors, *Proceedings of Graphics Interface*, pages 191–198, Mississauga, ON, Canada, 2007. Canadian Information Processing Society.

PETRA NEUMANN, ANTHONY TANG, and SHEELAGH CARPENDALE. A Framework for Visual Information Analysis. Technical Report 2007-87123, University of Calgary, Calgary, AB, Canada, 2007.

MEREDITH RINGEL MORRIS, KATHY RYALL, CHIA SHEN, CLIFTON FORLINES, and FREDERIC VERNIER. Beyond "Social Protocols": Multi-User Coordination Policies for Co-located Groupware. In *Proc. of Computer-Supported Cooperative Work (CSCW)*, pages 262–265, New York, USA, 2004. ACM Press.

GEORGE G. ROBERTSON, JOCK D. MACKINLAY, and STUART K. CARD. Cone Trees: Animated 3D Visualizations of Hierarchical Information. In *Conference Companion of the Conference of Human Factors in Computing Systems, CHI'91*, pages 189–194, New York, NY, USA, 1991. ACM Press.

DANIEL M. RUSSELL, MARK J. STEFIK, PETER PIROLLI, and STUART K. CARD. The Cost Structure of Sensemaking. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 269–276, New York, NY, USA, 1993. ACM Press.

STACEY D. SCOTT, M. SHEELAGH T. CARPENDALE, and KORI M. INKPEN. Territoriality in Collaborative Tabletop Workspaces. In *Proceedings of Computer-Supported*

Cooperative Work (CSCW), pages 294–303, New York, NY, USA, 2004. ACM Press.

BEN SHNEIDERMAN. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the Symposium on Visual Languages*, pages 336–343, Los Alamitos, USA, 1996. IEEE Comp. Society.

ROBERT SPENCE. A Framework for Navigation. *International Journal of Human-Computer Studies*, 51(5):919–945, 1999.

ROBERT SPENCE. *Information Visualization*. Pearson Education Limited, Harlow, England, 2nd edition, 2007.

SPOTFIRE, INC. Decisionsite® Posters. Website, 2007. http://www.spotfire .com/ (accessed March 2007).

JOHN STEWART, BENJAMIN B. BEDERSON, and ALLISON DRUIN. Single Display Groupware: A Model for Co-present Collaboration. In *Conference on Human Factors in Computing Systems (CHI, Pittsburgh, PA, USA)*, pages 286–293, New York, NY, USA, 1999. ACM Press.

ANTHONY TANG, MELANIE TORY, BARRY PO, PETRA NEUMANN, and SHEELAGH CARPENDALE. Collaborative Coupling over Tabletop Displays. In *Proceedings of Human Factors in Computing Systems (CHI)*, pages 1181–1290, New York, 2006. ACM Press.

JAMES J. THOMAS and KRISTIN A. COOK, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.

JACK J. VAN WIJK and WIM A. A. NUIJ. Smooth and Efficient Zooming and Panning. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '03)*, pages 15–22, Los Alamitos, CA, USA, 2003. IEEE Press.

JARKE J. VAN WIJK and EDWARD R. VAN SELOW. Cluster and Calendar based Visualization of Time Series Data. In GRAHAM WILLS and DANIEL KEIM, editors, *Proceedings of the IEEE Symposium on Information Visualization (Infovis '99)*, pages 4–9, Los Alamitos, CA, USA, 1999. IEEE Press.

FRED VERNIER, NEAL LESH, and CHIA SHEN. Visualization Techniques for Circular Tabletop Interfaces. In *Proceedings of Advanced Visual Interfaces (AVI)*, pages 257–263, New York, USA, 2002. ACM Press.

FERNANDA B. VIÉGAS, MARTIN WATTENBERG, FRANK VAN HAM, JESSE KRISS, and MATT MCKEON. Many Eyes: A Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2007)*, 12(5):1121–1128, 2007.

COLIN WARE. *Information Visualization – Perception for Design*. Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann Publishers, Amsterdam, Boston, Heidelberg, and others, 2. edition, 2000.

PIERRE WELLNER. Interacting with Paper on the DigitalDesk. *Communications of the ACM*, 36(7):87–96, 1993.

GEROLD WESCHE, JÜRGEN WIND, MARTIN GÖBE, LARRY ROSENBLUM, JIM DURBIN, ROBERT DOYLE, DAVID TATE, ROBERT KING, BERND FRÖHLICH, MARTIN FISCHER, MANEESH AGRAWALA, ANDREW BEERS, PAT HANRAHAN, and STEVE BRYSON. Application of the Responsive Workbench. *IEEE Computer Graphics and Applications*, 17 (4):10–15, 1997.

TORRE ZUK, LOTHAR SCHLESIER, PETRA NEUMANN, MARK S. HANCOCK, and M. SHEELAGH T. CARPENDALE. Heuristics for Information Visualization Evaluation. In *Proceedings of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV 2006)*, pages 55–60, New York, NY, USA, 2006. ACM Press.

Interactive Tree Comparison for Co-located Collaborative Information Visualization

Petra Neumann and Sheelagh Carpendale

Abstract—In many domains increased collaboration has lead to more innovation by fostering the sharing of knowledge, skills, and ideas. Shared analysis of information visualizations does not only lead to increased information processing power, but team members can also share, negotiate, and discuss their views and interpretations on a dataset and contribute unique perspectives on a given problem. Designing technologies to support collaboration around information visualizations poses special challenges and relatively few systems have been designed. We focus on supporting small groups collaborating around information visualizations in a co-located setting, using a shared interactive tabletop display. We introduce an analysis of challenges and requirements for the design of co-located collaborative information visualization systems. We then present a new system that facilitates hierarchical data comparison tasks for this type of collaborative work. Our system supports multi-user input, shared and individual views on the hierarchical data visualization, flexible use of representations, and flexible workspace organization to facilitate group work around visualizations.

Index Terms—Information visualization, collaboration, co-located work, hierarchical data comparison.

1 Introduction

It is common for small groups to gather around information that often involves some sort of visualization. Imagine a team of medical practitioners (doctors, nurses, physiotherapist, social workers) examining a patient's medical record to create a discharge plan, a team of geologists gathering around a large map to plan an upcoming expedition, or a team of executives looking at charts showing the latest sales trends. In many disciplines, collaboration allows for a multi-disciplinary group with an increased skill set. Different team members offer different perspectives and expertise that together can improve the quality of the solutions. Analyzing data collaboratively can also have several benefits. For instance, the information space may simply be too complex for an individual to interpret in its entirety, or the dataset may be susceptible to a variety of interpretations, in which case experts may discuss and negotiate their interpretations of the data. With large data sets, even the task load of *exploring* the data could be shared among several individuals on a team [33]. The benefits that collaboration offers to this process have motivated us to shift our efforts from single-user information visualization tools toward the design of *collaborative* information visualization tools.

Current information visualizations have mostly been designed from a single-user perspective. While it is possible for small teams to work with information visualizations using the standard setup of a small screen, one mouse and one keyboard, only one person at a time is able to make any changes to the view of the system. Attempting to collaborate under these conditions can be awkward and unnatural. The recent trend toward the use of large interactive displays offers the potential for the development of improved collaborative information visualization systems in which many co-located users can simultaneously interact and explore data sets. However, it is not yet understood how interfaces, visualizations, and interaction techniques should be designed to specifically address the needs of small co-located groups. The research problem we address is that, while most information visualization tools support sophisticated interaction with data, they have only limited facilities to support the collaborative activity of a team [14].

Research into supporting computer-supported cooperative work (CSCW) has considered both collaborations across distances (distributed) and collaborations where the team shares the same workspace (co-located) [12]. In our research, we focus on co-located collaboration. We present a collaborative information visualization system that supports collaborative analysis of data for small groups as they gather around an interactive digital table. Hierarchical data visualizations can be explored and compared in our environment through the use of several features that have been designed to facilitate collaborative work practices: multi-user input, shared and individual views on the hierarchical data visualization, flexible use of representations, and flexible workspace organization.

Our main contributions are: an analysis of challenges and requirements for the design of co-located collaborative information visualizations and a visualization system for collaborative tree comparison tasks around a large multi-touch tabletop display.

• *Petra Neumann is with the University of Calgary, E-mail: pneumann@cpsc.ucalgary.ca.*
• *Sheelagh Carpendale is with the University of Calgary, E-mail: sheelagh@cpsc.ucalgary.ca.*

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 2 November 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

2 Related Work

We start by reviewing the relatively small body of research that is directly concerned with the problem of supporting collaborative work around visual information. Then we consider research in the synchronous and asynchronous distributed domains, followed by a discussion of the work that focuses on supporting group analysis of visualizations in a shared space. Lastly, we discuss related work in the area of hierarchical data comparison to lay a foundation for our collaborative visualization system.

2.1 Collaborative Information Visualization

Research on the *process* of collaborative data analysis using information visualizations is relatively scarce. Mark *et al.* [14, 15] conducted a user study in which they observed pairs working in co-located and distributed settings with two different visualization systems designed for single users. Their findings suggest that the benefit of collaborative vs. individual problem solving was heavily dependent on the visualization system used and also that, in general, groups were better at locating errors [15]. In their second paper, Mark *et al.* [14] introduce a model for the collaborative problem-solving process. The model consists of an iterative sequence of five stages: parsing a question, mapping variables to the program, finding the correct visualization, and two validation stages. For collaborative work on scientific visualizations in virtual environments using CAVEs, Park *et al.* [20] report a five-step activity model that was common for the observed collaboration sessions. Their study also noted that participants showed a strong tendency for independent work, if the option was available.

2.1.1 Distributed Visualization Systems

In the area of scientific visualization, distributed systems have been introduced as early as 1994 [1] (see [6] for an overview). There is less research focused on distributed collaborative systems more directly concerned with *information visualization*. CoMotion is a collaborative environment for creating information analysis and decision-support applications [16]. The application provides shared views of the data on which all users can *synchronously* interact. The Command Post of the Future is a visualization tool built on this architecture in which distributed team members can share visualizations of natural emergencies and combat situations [3]. Examples of web-based *asynchronous* collaborative environments include sense.us [9], Many Eyes[1], and Swivel.[2] They all allow collaborative sharing and exploration of data by letting users upload and visualize data as well as comment on created visualizations. DecisionSite Posters [17], another web-based system for asynchronous collaborative work around information visualization, allows users to publish visualization results, descriptions, and data to distributed collaborators in interactive web-based reports.

2.1.2 Co-located Visualization Systems

The responsive workbench was one of the first visualization systems for co-located collaboration around a large horizontal surface [37]. The responsive workbench is a virtual reality environment in which the displayed 3D scene is looked at through shuttered glasses. Several scientific visualization applications were developed for this platform including fluid dynamics and situational awareness applications.

With a focus on interaction, radial tree layouts have been studied for collaborative circular tabletop systems [35]. In their system Vernier *et al.* provide two different fisheye mechanisms to support different types of user activity around the tabletop. Interactive focus+context techniques for collaborative systems have further been applied in the DTLens system [4]. In general, no guidelines, as of yet, exist for the development of collaborative systems specifically tailored for information visualization applications.

2.2 Hierarchical Data Comparison Systems

The work most closely related to our approach is the TreeJuxtaposer system by Munzner *et al.* [18]. In this work structural comparison of nodes is facilitated by finding the most similar (or best corresponding node) to one tree in another tree. The best corresponding node(s) and dissimilar nodes are highlighted in their system giving an overview of structural differences and similarities between trees. Graham and Kennedy [5] present a system for linked highlighting across several hierarchies. Similarities are shown by giving similar nodes new degree-of-interest values leading to a more prominent display in the hierarchy. We chose to use the similarity measure as described in [18] to calculate similarities across our hierarchies. Our system extends comparative possibilities by incorporating collaborative comparative interactions.

3 Design Guidelines for Co-located Collaborative Information Visualization Systems

In this section we discuss design guidelines specifically for *co-located collaborative information visualization systems*. These design heuristics are condensed from information visualization design advice [29, 38, 39, 40], co-located collaboration advice [8, 13, 23, 21, 24, 25, 26, 27, 28, 32], the studies that look directly at collaborative visualization [14, 15, 20] and our observations of teams of people collaborating to solve tasks using information visualizations [19]. Other design heuristics exist that can guide the developer of an information visualization system in terms of the data domain, cognitive levels based on knowledge and task, or perception and cognition [41]. In the field of computer supported collaborative work (CSCW) a set of main design guidelines has been established for collaborative systems in general (e. g., [21, 28]); however, they do not take the specific problems and requirements of information visualization applications into account. Our intention is that these design guidelines, compiled from three bodies of research, will form a basis which will adjust and expand as research in collaboration around information visualizations continues.

[1]http://services.alphaworks.ibm.com/manyeyes/home/

[2]http://www.swivel.com/

3.1 Hardware and System Setup

In this section we consider guidelines for the design of the physical workspace in which the collaborative activities around information visualization can take place.

Size: In information visualization, the size of the available display space has always been problematic for the representation of large datasets (e. g., [18]). In a common desktop environment, typically a single user will use all available screen space to display their visualization and, most commonly, this space will not be sufficient. Frequently, visualization software will include interactive features to help the user cope with limited display space. It seems sensible to think that, if we are going to adequately support collaborative or team exploration of visualizations, available display space will be an important issue. In collaborative systems, screen space has not only to be large enough for the required information display, it might also have to be viewed and shared by several users. As the number of people using a shared information display grows, *the size of the display and workspace needs to be increased* in order to provide a large enough viewing and interaction area that gives equal access to all group members.

Configuration: Several configuration possibilities exist that could enlarge an information display, all of which will affect the type of visualization system possible and the type of collaboration work that would be most readily supported. Many types of configurations are possible, for instance one could provide team members with interconnected individual displays, as in the ConnecTable system [31], or one could make use of large interactive single-display technology, like display walls or interactive tabletop displays (e. g., [32]). An additional possibility is to link wall, table, and personal displays (e. g., [38]), or to consider immersive displays (e. g., [20]). The type of setup most appropriate for an information visualization system will depend on the specific task and group setup. For example, individual interconnected displays allow for private views of at least parts of the data which might be required if data access is restricted. Tabletop displays have been found to encourage group members to work together in more cohesive ways, whereas wall displays are beneficial if information has to be discussed with a larger group of people [24].

Input: In the common desktop setup, input is provided for one person through keyboard and mouse. To support collaboration, ideally, each person would have at least one means of input. In addition, it would be helpful if this input was identifiable, making it possible to personalize system responses. If a collaborative system allows for multi-user input, the access to a shared visualization and data set has to be coordinated. Also, synchronous interactions on a single representation may require the design and implementation of new types of multi-focus visualizations. Ryall *et al.* [25] have addressed the problem of personalization of parameter changes for widget design. Based on user identity, their widgets can be dynamically adapted for individuals within a group. Similar ideas could be implemented for personalization of information visualizations during collaborative work.

Resolution: Resolution is an issue both for the output (the display) and for the input. The display resolution also has a great influence on the legibility of information visualizations. Large display technology currently often suffers from relatively low display resolution so that visualizations might have to be re-designed if readability of text, color, and size is affected by display resolution. Large interactive displays are often operated using fingers or pens which have a rather low input resolution. Since information visualizations often display large data sets with many relatively small items, the question of how to select these small items using low input resolution techniques becomes an additional challenge that needs special attention [11].

Interactive response: Implementations of collaborative information visualizations have to be carefully designed for efficiency. Individual information displays can already be computationally intensive and require considerable pre-processing (e. g., [18]). Yet, in collaborative systems several information visualizations might have to be displayed and interacted with at the same time. While powerful hardware can solve the problem to some extent, efficient data processing as well as fast rendering of the graphical representations should be considered when the needs of several users have to be addressed.

3.2 Designing the Information Visualization

Many known information visualization guidelines will still apply for the design of information visualizations for large displays or collaborative use (e. g., [2, 34, 36]). In this section we discuss additional aspects that need to be considered for the design of information visualizations for collaborative settings.

Supporting Mental Models: It has been shown that providing capabilities to freely move interface items is critical for group interactions and task coordination during co-located collaborative work [26]. Letting users impose their own organization on items in the workspace may help collaborators *create and maintain mental models* of a dataset that contains several different representations. By freely moving representations, team members can impose their own *categorizations* on the representations by, for example, placing them in close proximity or in piles relevant to a task.

Representation Changes: Zhang and Norman [40] found that providing different representations of the same information to individuals provides different task efficiencies, task complexities, and changes decision-making strategies. In a collaborative situation, group members might have different preferences or conventions that favour a certain type of representation. Gutwin and Greenberg [8] have discussed how different representations of the workspace affect group work in a distributed setting. They point out that providing multiple representations can aid the individual but can restrict how the group can communicate about the objects in the workspace. This extends to co-located settings in which several representations of a dataset can be personalized according to taste or convention making it harder to relate individual data items in one representation to a specific data item in another. For example, relating one specific node in a Treemap to another node in a node-link diagram might require a search to locate the respective node in the other representation. Implementing mechanisms to highlight individual data items across representations might aid individuals when switching between group and more parallel data exploration.

Task History: Collaborative information visualization systems should also provide access to some form of data analysis history. While this is true for information visualizations in general [29], it might be of even higher importance in collaborative settings. Chua and Roth [3] have suggested that capturing and visualizing information about interactions of collaborators with objects in a workspace may enhance collaboration by leading to a better understanding of each others' involvement in solving a task. As group members switch between work on individual and shared views of the data, they might lose track of the interactions of their collaborators [8]. The access to an exploration history can help in later discussing the data and exploration results with collaborators or informing them about interesting data aspects that have been found during the analysis process.

Perception: Relatively little has been done to analyze how the interpretation of information visualizations is affected when viewed on different display configurations. A study by Yost and North [39] evaluated the scalability of three visualizations across a small and large, high-resolution display. Their study does not take the requirements of collaboration into account but provides several guidelines for designing visualizations for large displays: considering encodings according to viewing angle, choosing visualizations for scalable encoding, providing global and local legends, and strategic label placement. A study by Wigdor *et al.* [38] evaluated the effect of viewing angle on different graphical variables and suggests that care should be taken in positioning and choosing the appropriate visual encoding as some graphical elements are more robust to distortion than others. In the case of collaborative work around a large horizontal display, group members might be positioned on different sides of the display, thus viewing shared visualizations from different directions. It has to be evaluated how the legibility of information visualizations is affected by different viewing directions. So far, it is not known if, for example, an upside-down bar chart would lead to inaccurate readings of the data.

3.3 Designing the Collaborative Environment

Pinelle *et al.* [21] provide a set of basic operations that should be supported by groupware systems to help collaborators carry out their tasks as a team. These mechanics of collaboration can be grouped into those describing communication and coordination aspects of collaboration. Collaborative information visualization systems also require additional support for communication, coordination, and changing collaboration styles to further the analytics process.

Coordination: In group settings, collaborators have to coordinate their actions with each other. Here, we describe several guidelines for how to support the coordination of activities in collaborative information visualization applications.

Workspace Organization: Typical single-user information visualization systems impose a fixed layout of windows and controls in the workspace. Previous research has shown that, on shared workspaces, collaborators tend to divide their work areas into personal, group, and storage territories [27]. This finding implies that a *group interaction and viewing space* is needed for collaborative data analysis where the group works on a shared representation of the data or in which shared tools and representations. Also, the possibility to explore the data separately from others, in a *personal space*, is necessary.

Fluid Interaction: Collaborative systems should support fluid transitions between activities to improve the coordination of activities [28]. This implies that information analysis tasks that require the application of tools (filters, lenses, ...) or changing of view or visualization parameters should be designed to require (a) as little shift of input mode (mouse, keyboard, pen, finger, ...) as possible, and (b) as little manipulation of interface widgets and dialogs as possible. For information visualizations, this is a difficult design problem, as systems frequently offer extensive lists of parameters to manipulate in order to provide flexible interaction. Similarly, the study on collaborative information visualization by Mark *et al.* [15] suggests that groups work more effectively if the interactions with a system are easier to understand.

Information Access: Information access through information visualizations also needs to be coordinated on a global and local scope. What if one group member found something in the data that he or she wishes to delete or modify? Who can change the scale, zoom, or rotation settings for a shared view of the data? Policies might have to be put in place to restrict certain members from making unsuspected global changes to the data that might change other group members' view of the same data [23].

Collaboration Styles: Tang *et al.* [32] describe how collaborators tend to frequently switch between different types of loosely and closely coupled work styles when working over a single, large, spatially-fixed information display (e. g., maps or network graphs). A study by Park *et al.* [20] in distributed CAVE environments discovered that, if the visualization system supports an individual work style, users preferred to work individually on at least parts of the problem. For information visualization systems, an individual work style can be supported by providing *access to several copies of one representation*. The availability of unlimited copies of one type of representation of data allows group members to work in parallel. More closely coupled or joint work on a single view of the data can be supported by implementing the possibility of *concurrent access and interaction* with the parameters of an information visualization. Free arrangements of representations also *supports changing work styles*. Representations can be fluidly dragged into personal work areas for individual or parallel work and into a group space for closer collaboration.

Communication: Communication is an important part of successful collaborations. People need to be able to trigger conversations, communicate about intentions to change collaboration styles, indicate a need to share a visualization, and to be generally aware of their team members' actions. Providing awareness of global changes is important to support communication about the information analysis process [8, 3]. Group members need to be informed that some parameter of a shared display might have changed while they were busy working with an information visualization in a different part of the workspace. If group members decided to work in parallel on different subproblems, the visual comparison of the individual graphical exploration results has to be supported in order to make group discourse on the results possible. To enrich the discourse about individual visualization exploration results, additional interaction schemes such as annotation of the results should also be included [9].

Flexible workspace organization can offer the benefit of easy sharing, gathering, and passing of representations to other collaborators. By sharing data in the workspace, representations will be viewed by team members with possibly different skill sets and experiences and, therefore, subjected to different interpretations. Also, by being able to move and rotate representations in the workspace, an individual can gain a new view of the data and maybe discover previously overlooked aspects of the data display. Communication can also be supported through the design of gathering and sharing mechanisms. However, the design of these mechanisms needs to respect common social and work protocols [13, 23, 27]. For example, the interface should not require a group member to reach into or across another person's workspace in order to acquire or share visualizations or controls.

4 A System for Co-located Collaborative Work with Information Visualizations

This section provides a detailed description of our information visualization system designed to support collaborative tree comparison tasks. Paralleling our design guidelines section we describe our hardware setup, our information visualizations and then those aspects specially included to support collaboration. Then, in Section 5, we describe this system in use for a collaborative tree comparison task.

4.1 Hardware and System Setup

Our system was designed to run on a large digital tabletop display; however, using it on large wall displays is also possible. Our digital table is built using a touch-sensitive DViT Board from SMART Technologies with two concurrent and independent inputs (see Figure 1). The tabletop setup has $2{,}800 \times 2{,}100$ pixels (≈ 5.9 mega pixels) provided by four rear-mounted projectors (2×2). This setup offers an adequate size, configuration, input, and resolution for small groups of 2–4 individuals to work together. However, only two simultaneous touches are currently supported by our technology and inputs are not identifiable. Our implementation is based on a general framework for tabletop interfaces that provides a method of spatially representing properties of the interface using a buffer approach [10]. This framework and the buffer approach are able to maintain interactive response on high-resolution tabletop displays. We use the framework, for example, to implement picking and interaction regions for our widgets. The framework also provides access to other tabletop interaction metaphors and widgets such as RNT [13], tossing, and Storage Territories [26]. To facilitate not only an efficient management of memory resources but also to allow people to relate one visual representation of a dataset to a different one of the same data, we maintain only one copy of this underlying dataset. Each visual representation of a dataset is then realized using a set of meta data to represent the specific visual appearance.

4.2 Information Visualizations

Our system supports work with hierarchical data, specifically with two different types of tree representations: a space-filling radial tree layout and a cladogram. We have chosen to implement a radial tree layout as presented in [30], with a minor adjustment that places labels in a circular fashion inside the nodes (see Figure 2, left). We chose this type of labeling to facilitate orientation-independent reading from different positions around the tabletop display. Since tree comparison is a task commonly performed on phylogenetic trees [18] we also implemented a cladogram tree layout (see Figure 2, right). In the cladogram layout, all leaf nodes are extended to the bottom of the graph. To additionally reveal their place in the hierarchy, nodes are coloured according to their level. Our system can easily be extended to support other types of representations.

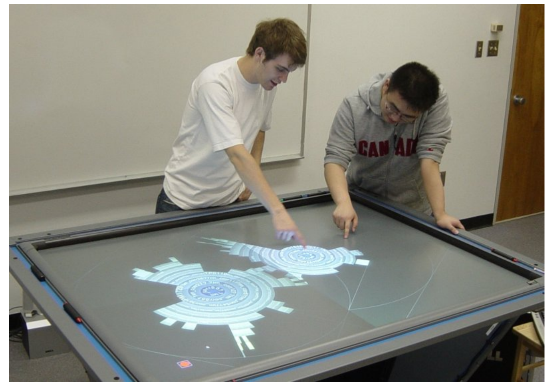


Fig. 1. The hardware setup for our collaborative information visualization application. Two simultaneous pen or finger inputs are possible.

Any information visualization and all control widgets in our system can be freely re-oriented and repositioned. Each information visualization is drawn on its own plane with appropriate controls attached to the side. The left of Figure 3 shows a single visualization plane showing a radial tree layout and its attached menu buttons. The menu offers common view parameter changes: scaling (zoom), integrated rotation and translation [13], translation only, and annotation. Thus, the plane and attached visualization can be freely moved around the tabletop display. The right of Figure 3 shows an arrangement of three visualization planes on the tabletop display.

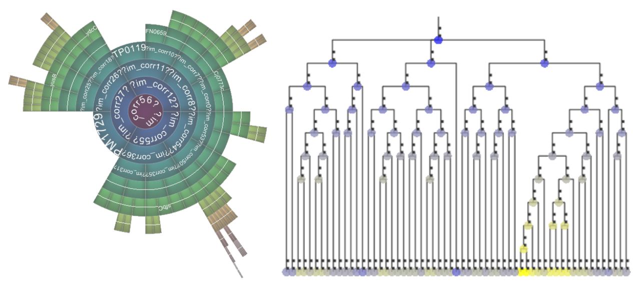


Fig. 2. The two representations used in our system. *Left:* a radial tree layout with radial labeling. *Right:* a cladogram with additional node colouring to reveal level information.

4.2.1 Supporting Mental Models

In Section 3, we have identified the creation and maintenance of mental models of the data set as one of the possible benefits of allowing users, rather than the interface designer, to impose a layout of visualizations in the workspace. By supporting free rotation, translation, and scale, users of our system can create their own organization of items by putting them in piles creating a preferred layout (e. g., small multiples). The possibility for organizing representations of data is further supported by providing storage containers that hold visualization planes. In these containers, visualizations can be grouped together, resized, and moved as a unit (as in [26]). Figure 4 shows an example of a visualization plane being placed in a storage container. First, the

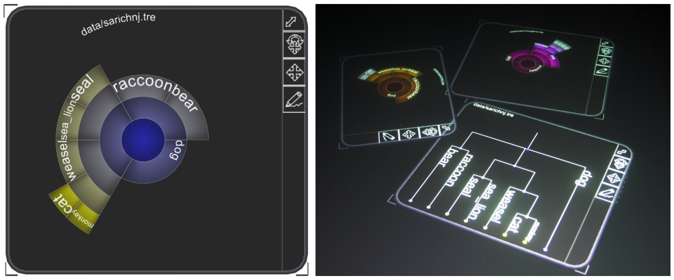


Fig. 3. A single *visualization plane* showing a radial tree layout can be seen on the left. The right image shows three visualization planes oriented on the tabletop display.

plane is dropped on the container (left), and then automatically resized and placed in the storage container (right). Items in the storage container can be placed casually, neatly organized, or piled, and can then be moved as a unit. These containers can provide a means for collaborators to store intermediate exploration results for later reference or comparison.

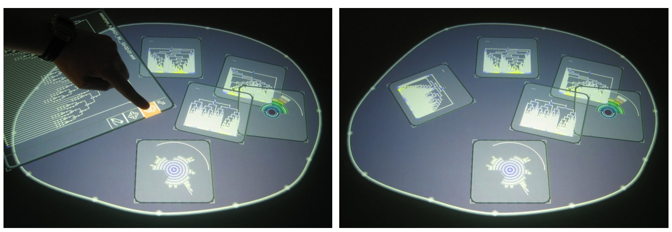


Fig. 4. A visualization plane is being dropped on a storage container (left) and automatically resized and placed (right).

4.2.2 Representation Changes

To support changing decision-making strategies and personal tastes and conventions, we provide individual access to different types of representations. If an individual group member wishes to visualize the data using a different but appropriate representation of the data, e. g., a containment tree layout instead of a node-link diagram, the specific representation can be changed with a drag-and-drop operation without interfering with other group members' operations. Figure 5 shows how a representation change is performed. In the left image the visualization plane is dragged onto the RepresentationChanger widget. As soon as the plane has been placed on the widget, the representation changes to the desired one as can be seen at the right of Figure 5.

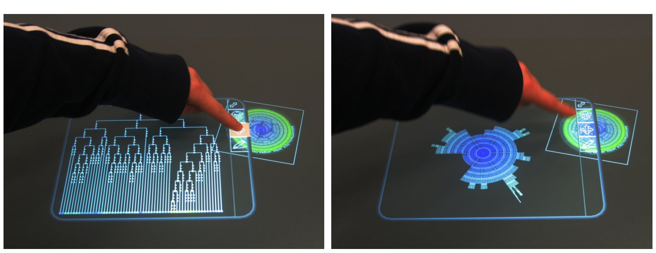


Fig. 5. Switching a representation type with a drag-and-drop operation.

4.2.3 Task History and Perception

Our system currently only includes annotation and note taking capabilities to capture exploration history (see Section 4.3.3). Further capabilities will be designed and evaluated for future versions of our system. As few evaluations (e. g. [39, 38]) have discussed the effects of perspective distortion and orientation on the readability of information visualizations we have not attempted to correct for possible negative effects.

4.3 Design for Collaboration

In this section we discuss the features of our system according to guidelines for the design of the collaborative environment as presented in Section 3. At this time we have addressed issues concerning workspace organization, fluid interaction, supporting differing collaboration styles, and communication.

4.3.1 Workspace Organization and Collaboration Styles

Free workspace organization allows us to support different work styles. Collaborators can fluidly transition between more independent work and closer, joint work on information visualizations. Figure 6 gives an example in which two collaborators are working individually at first, looking at visualizations in their own area of the workspace (Figure 6, left) and then switch to a more closely coupled work style by investigating one visualization together in more detail (Figure 6, right). Note that the scaling mechanism has been applied to create a larger visualization to accommodate the concurrent interaction and viewing of both partners and that the plane has been rotated towards both team members. This type of rotation has been previously identified as a strong communicative gesture [13].

Fig. 6. Visualization planes can be freely arranged in our system. On the left two collaborators are looking at a few representations individually. On the right they are investigating one visualization together.

Any number of windows can be created, moved, and interacted with in the workspace, limited only by the complexity of the graphics and the capabilities of the graphics hardware. By allowing collaborators to each access a copy of a representation we support parallel work on the same data. Each dataset loaded into our system is represented as a floating menu entry in the workspace, as can be seen at the left of Figure 7. The menu entries can be freely repositioned and, thus, passed to other collaborators to facilitate shared access to this resource. By pointing on the menu entry a new visualization plane with a representation of the data is created (see Figure 7, right). While initial response has been enthusiastic, we realize that many careful studies are required to evaluate the varying effects of our tools on group work.

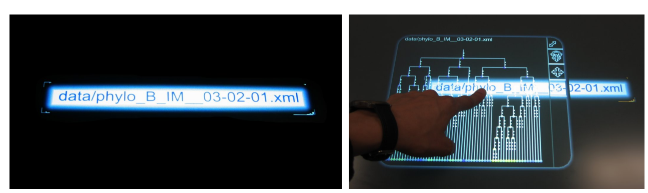


Fig. 7. Creation of additional representations using floating menu entries. *Left:* an example of a floating menu entry on the tabletop display. *Right:* A user created a new visualization by touching the menu entry.

4.3.2 Fluid Interaction

Our hardware supports input using fingers or pens and reports touch information (touch down, touch up) to the interface application. We, therefore, have to design interaction techniques without common interactions known from the desktop, like double, left, or right click.

In desktop software, dialog boxes or spatially fixed menus or parameter panels are usually used to change visualization parameters for a given focus view. Most traditional widgets are not suitable for providing fluid interaction (see Section 3) and to be concurrently accessible by several group members. During parallel work in a group setting several visualizations might have a focus at the same time or a visualization might be interacted upon by more than one person at a time. Research on a system for collective co-located annotation of digital photos revealed that users strongly preferred a replicated set of controls over a centralized shared set of controls because the center of the table was needed for other tasks and because replicated controls avoided accidental touching by other teammates [22]. We, therefore, opted for a replicated set of controls where each control could also be freely positioned.

Currently, we implemented visualization change parameters as drag and drop operations. For example, we implemented ColourChanger widgets on which a visualization can be dropped in order to initiate a change of its colour scale (see Figure 8). Alternatively, these widgets could also be dropped on the visualization plane in order to initiate a parameter change. This alternative would avoid having to reposition visualization planes if a careful layout has been created by the group. We are also experimenting with other input techniques like flow menus [7] in order to make a large number of parameters accessible for each visualization plane.

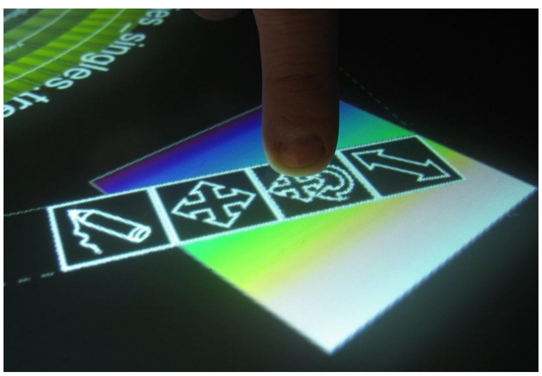


Fig. 8. A visualization plane is dropped on a ColourChanger widget that changes the colour scale with which the tree is displayed.

4.3.3 Communication

To enrich the discourse of individual and group exploration of data our system allows for annotation directly on the provided visualizations and separately on sticky notes. Interactive sticky notes for low-resolution input [11] can be used to take general notes during the exploration process to, for example, write down intermediate results or variables to look for. Using these annotations, collaborators can become aware of each others' exploration processes even if the individual work takes place in separate areas of the workspace. Figure 9 shows how sticky notes and integrated annotations can be used to mark interesting information in a tree layout. By allowing visualizations to be freely repositioned we offer a mechanism for *sharing of visualizations* as the windows can be be easily passed to the other collaborators. Representations can also be passed by dragging or tossing them across the table, similar as implemented for pictures in [26].

5 Collaborative Tree Comparison

In this section we show how our system functions by stepping through a task of collaborative tree comparison. We use an example based on the InfoVis 2003 contest[3] dataset, showing how our system supports collaborative comparison tasks.

[3]http://www.cs.umd.edu/hcil/iv03contest/

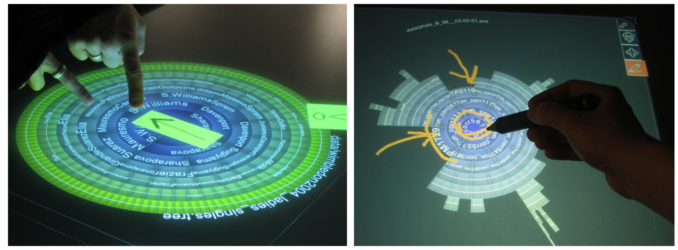


Fig. 9. Annotation of visualizations. *Left:* Annotation using interactive sticky notes [11]. *Right:* Annotation integrated directly on the information visualization.

5.1 Data and Task

As example data for our comparison tasks we used the InfoVis 2003 phylogenetic data and tasks. This dataset contains information on the evolution of two proteins (Protein ABC and Protein IM). It has been suggested that both proteins co-evolve and that such a co-evolution can be detected by comparing the phylogenies of both proteins. The high-level task was to find out whether such a co-evolution was visible. Lower-level comparison tasks included finding where structural changes occurred in the tree. We chose to use the two main files for the ABC and IM proteins and the additional four trees that were provided. We did not pair proteins between the two trees.

5.2 Tree Comparison Algorithm and Visualization

We used the same similarity measure as used for the TreeJuxtaposer system [18], which is based on comparing the sets of labels of nodes in the subtree under each node. The best corresponding node(s) and nodes with no similarity were highlighted. Figure 10 shows a comparison of two trees containing different versions of a carnivore hierarchy. The node "dog" has been selected by a user in the left tree. The best corresponding node "dog" in the right tree is highlighted in yellow, whereas nodes with no similarity are highlighted in red. Nodes in blue are not highlighted in the right tree as they contain the node "dog" (yellow) in their subtree and are therefore "somewhat similar."

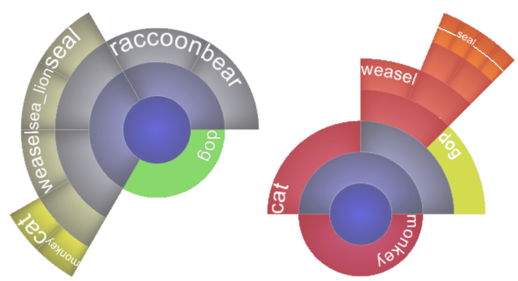


Fig. 10. Tree Comparison of two different versions of a carnivore data set. *Left:* The node "dog" has been selected for comparison. *Right:* The node "dog" is highlighted in yellow as the best corresponding node. Nodes in red have no correspondence with the node "dog."

Trees in our system can be compared by moving their visualization planes close to one another. When planes are close enough for comparison the borders are highlighted and nodes can be selected to start a similarity calculation. In Figure 11, we show two planes on the left in comparison mode (orange border) and a smaller tree to the side that is not currently compared. Any number of trees can be compared by moving them close to others that are already being compared.

5.3 Solving Collaborative Tree Comparison Tasks

To gain an overview of the available information, each visualization plane can be arranged to facilitate a comparison between all available datasets. In Figure 12 two users of our system created a comparison overview by organizing their planes to facilitate cross-comparison.

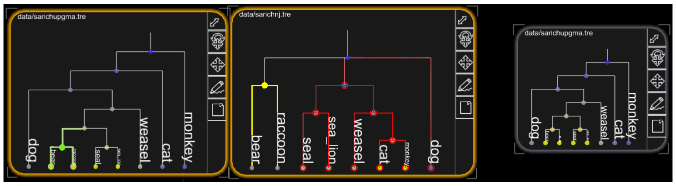


Fig. 11. Trees can be compared when their planes are in close proximity. Here the two planes on the left are in comparison mode as can be seen by the highlighted (orange) border. The tree on the right is not currently compared with the others.

Figure 13 shows a close-up screenshot of such a comparison. The middle two planes show the main IM and ABC protein representation. The root node of the ABC protein (top row) has been highlighted (green). The two trees on the left, the alternative versions of the IM protein, and the IM protein tree show only dissimilar nodes to the ABC protein (in red). However, the alternative versions of the ABC proteins both show a few dissimilar nodes that need to be inspected further.

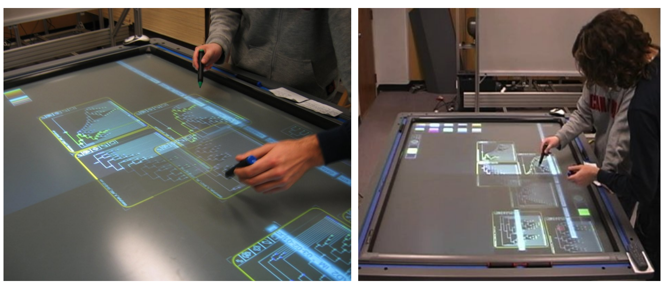


Fig. 12. All six datasets have been moved together to facilitate a comparison across all representations.

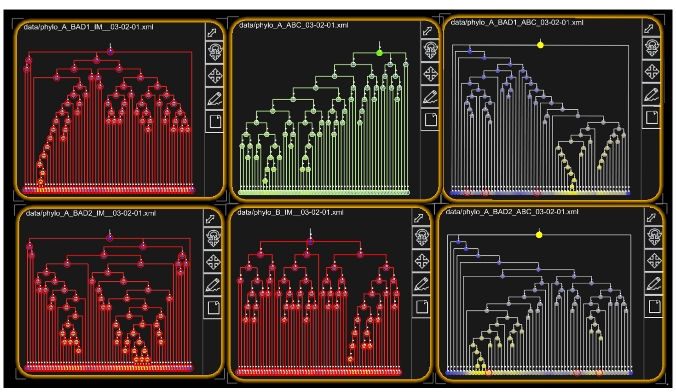


Fig. 13. Screenshot of the system showing all six trees. The root node of the ABC protein in the top center plane has been highlighted.

This more detailed investigation within the versions of the ABC and IM protein was performed in parallel. The left of Figure 14 shows two collaborators who have decided to each investigate one of the proteins. To inspect which nodes have dissimilar values, they have chosen to annotate the dissimilar nodes first and to then examine the nodes and their structure in the hierarchy in more detail. However, closer examination of nodes can also be performed in joint work as shown in Figure 14 (right).

A contest task required the examination of the hierarchical structure in terms of whether subtrees moved in the hierarchies or nodes changed position. To facilitate a structural comparison of nodes in this sense, trees in our system can be overlaid and then examined. All visualization planes are semi-transparent in order to support this type of tree comparison. Figure 15 gives two examples of structural comparison through overlay. The top image shows an overlay of Protein ABC (blue) and Protein IM (magenta). It can be seen that Protein ABC is generally more shallow than Protein IM but has one main subtree that is wider and deeper than can be found in the other tree. In the bottom image, two users overlaid their exploration history including annotations of similar trees. Similar and dissimilar nodes are highlighted. We are considering options to auto-rotate planes to show the best possible match.

Fig. 14. Closer examination of a few trees. *Left:* Parallel work with each person comparing three trees each. *Right:* Joint work comparing four trees together.

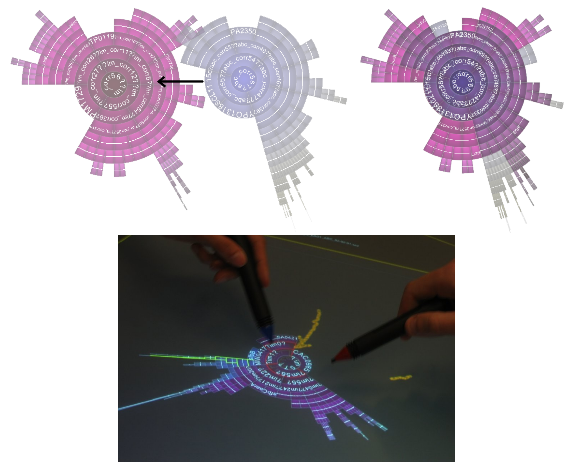


Fig. 15. Structural comparison through overlay.

6 Conclusion

In this paper we have provided guidelines for the design of co-located collaborative information visualization systems. We have applied these guidelines in the creation of a system for collaborative tree comparison tasks and have shown how such tasks can be solved in a collaborative fashion using our system. So far, most information visualization systems have been designed with a single user in mind. How, or whether, interfaces, visualizations, and interaction techniques should be designed to specifically address the needs and requirements for teams of individuals analyzing data still needs to be further explored. In this paper we have contributed to the evolving knowledge about the design of such systems. Our guidelines have been derived from general information visualization design advice, co-located collaboration advice, the few studies that look directly at collaborative visualization, and our observations of teams of people working together to solve tasks using information visualizations. As our collaborative system is evaluated and developed further, and as other researchers contribute to the development of collaborative information visualization systems, we expect these guidelines to be extended.

Acknowledgements

We would like to thank Tobias Isenberg for his work on the underlying tabletop framework and Anthony Tang, Mark S. Hancock, and many ilab members for giving comments on the paper and testing the system. We would also like to thank our funding agencies Alberta Ingenuity, iCORE, and NSERC.

References

[1] V. Anupam, C. Bajaj, D. Schikore, and M. Schikore. Distributed and Collaborative Visualization. *Computer*, 27(7):37–43, July 1994.

[2] J. Bertin. *Semiology of Graphics: Diagrams Networks Maps*. The University of Wisconsin Press, Madison, USA, 1983.

[3] M. C. Chuah and S. F. Roth. Visualizing Common Ground. In *Proc. of the Conf. on Information Visualization (IV)*, pages 365–372, Los Alamitos, USA, 2003. IEEE Comp. Society.

[4] C. Forlines and C. Shen. DTLens: Multi-user Tabletop Spatial Data Exploration. In *Proc. of User Interface Software and Technology (UIST)*, pages 119–122, New York, USA, 2005. ACM Press.

[5] M. Graham and J. Kennedy. Combining Linking & Focusing Techniques for a Multiple Hierarchy Visualisation. In *Proc. of the Conf. on Information Visualization (IV)*, pages 425–432, Los Alamitos, USA, 2001. IEEE Comp. Society.

[6] I. J. Grimstead, D. W. Walker, and N. J. Avis. Collaborative Visualization: A Review and Taxonomy. In *Proc. of the Symp. on Distributed Simulation and Real-Time Applications*, pages 61–69, Los Alamitos, USA, 2005. IEEE Comp. Society.

[7] F. Guimbretiére and T. Winograd. FlowMenu: Combining Command, Text, and Data Entry. In *Proc. of User Interface Software and Technology (UIST)*, pages 213–216, New York, NY, USA, 2000. ACM Press.

[8] C. Gutwin and S. Greenberg. Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. In *Proc. of Computer Supported Cooperative Work (CSCW)*, pages 207–216, New York, USA, 1998. ACM Press.

[9] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. In *Proc. of Human Factors in Computing Systems (CHI)*, pages 1029–1038, New York, USA, 2007. ACM Press.

[10] T. Isenberg, A. Miede, and S. Carpendale. A Buffer Framework for Supporting Responsive Interaction in Information Visualization Interfaces. In *Proc. of* C^5, pages 262–269, Los Alamitos, USA, 2006. IEEE Comp. Society.

[11] T. Isenberg, P. Neumann, S. Carpendale, S. Nix, and S. Greenberg. Interactive Annotations on Large, High-Resolution Information Displays. In *Conf. Comp. of IEEE VIS, InfoVis, and VAST*, pages 124–125, Los Alamitos, 2006. IEEE Comp. Society.

[12] R. Johansen. *Groupware—-Computer Support for Business Teams*. The Free Press, New York, USA, 1988.

[13] R. Kruger, S. Carpendale, S. D. Scott, and S. Greenberg. Roles of Orientation in Tabletop Collaboration: Comprehension, Coordination and Communication. *Journal of Computer Supported Collaborative Work*, 13(5–6):501–537, 2004.

[14] G. Mark, K. Carpenter, and A. Kobsa. A Model of Synchronous Collaborative Information Visualization. In *Proc. of the Conf. on Information Visualization (IV)*, pages 373–381, Los Alamitos, USA, 2003. IEEE Comp. Society.

[15] G. Mark, A. Kobsa, and V. Gonzalez. Do Four Eyes See Better than Two? Collaborative versus Individual Discovery in Data Visualization Systems. In *Proc. of the Conf. on Information Visualization (IV)*, pages 249–255, Los Alamitos, USA, 2002. IEEE Comp. Society.

[16] MayaViz. Comotion®. Website, 2007. http://www.mayaviz.com/ (accessed March 2007).

[17] Spotfire, Inc. Decisionsite® Posters. Website, 2007. http://www.spotfire.com/ (accessed March 2007).

[18] T. Munzner, F. Guimbretire, S. Tasiran, L. Zhang, and Y. Zhou. TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context with Guaranteed Visibility. *ACM Transactions on Graphics*, 22(3):453–462, July 2003.

[19] P. Neumann, A. Tang, and S. Carpendale. A Framework for Visual Information Analysis. Technical Report 2007-87123, University of Calgary, Calgary, AB, Canada, July 2007.

[20] K. S. Park, A. Kapoor, and J. Leigh. Lessons Learned from Employing Multiple Perspectives In a Collaborative Virtual Environment for Visualizing Scientific Data. In *Proc. of Collaborative Virtual Environments (CVE)*, pages 73–82, New York, USA, 2000. ACM Press.

[21] D. Pinelle, C. Gutwin, and S. Greenberg. Task Analysis for Groupware Usability Evaluation: Modeling Shared-Workspace Tasks with the Mechanics of Collaboration. *ACM Transaction of Human Computer Interaction*, 10(4):281–311, Dec. 2003.

[22] M. Ringel Morris, A. Paepcke, T. Winograd, and J. Stamberger. TeamTag: Exploring Centralized versus Replicated Controls for Co-located Tabletop Groupware. In *Proc. of Human Factors in Computing Systems (CHI)*, pages 1273–1282, New York, USA, 2006. ACM Press.

[23] M. Ringel Morris, K. Ryall, C. Shen, C. Forlines, and F. Vernier. Beyond ”Social Protocols”: Multi-User Coordination Policies for Co-located Groupware. In *Proc. of Computer-Supported Cooperative Work (CSCW)*, pages 262–265, New York, USA, 2004. ACM Press.

[24] Y. Rogers and S. Lindley. Collaborating Around Vertical and Horizontal Large Interactive Displays: Which Way is Best? *Interacting with Computers*, 16(6):1133–1152, Dec. 2004.

[25] K. Ryall, A. Esenther, C. Forlines, C. Shen, S. Shipman, M. Ringel Morris, K. Everitt, and F. Vernier. Identity-Differentiating Widgets for Multiuser Interactive Surfaces. *IEEE Computer Graphics and Applications*, 26(5):56–64, Sept./Oct. 2006.

[26] S. D. Scott, M. S. T. Carpendale, and S. Habelski. Storage Bins: Mobile Storage for Collaborative Tabletop Displays. *IEEE Computer Graphics and Applications*, 25(4):58–65, July/Aug. 2005.

[27] S. D. Scott, M. S. T. Carpendale, and K. M. Inkpen. Territoriality in Collaborative Tabletop Workspaces. In *Proc. of Computer-Supported Cooperative Work (CSCW)*, pages 294–303, New York, NY, USA, 2004. ACM Press.

[28] S. D. Scott, K. D. Grant, and R. L. Mandryk. System Guidelines for Co-located Collaborative Work on a Tabletop Display. In *Proc. of the European Conf. on Computer-Supported Cooperative Work (ECSCW)*, pages 159–178, Dordrecht, The Netherlands, 2003. Kluwer Academic Publishers.

[29] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proc. of the IEEE Symp. on Visual Languages*, pages 336–343, Los Alamitos, USA, 1996. IEEE Comp. Society.

[30] J. T. Stasko and E. Zhang. Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations. In *Proc. of the Symp. on Information Visualization (InfoVis*, pages 57–65, Los Alamitos, USA, 2000. IEEE Comp. Society.

[31] P. Tandler, T. Prante, C. Müller-Tomfelde, B. Streitz, and R. Steinmetz. ConnecTables: Dynamic Coupling of Displays for the Flexible Creation of Shared Workspaces. In *Proc. of User Interface Software and Technology (UIST)*, pages 11–20, New York, USA, 2001. ACM Press.

[32] A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative Coupling over Tabletop Displays. In *Proc. of Human Factors in Computing Systems (CHI)*, pages 1181–1290, New York, 2006. ACM Press.

[33] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, Aug. 2005.

[34] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphic Press, Cheshire, Connecticut, USA, 2001.

[35] F. Vernier, N. Lesh, and C. Shen. Visualization Techniques for Circular Tabletop Interfaces. In *Proc. of Advanced Visual Interfaces (AVI)*, pages 257–263, New York, USA, 2002. ACM Press.

[36] C. Ware. *Information Visualization – Perception for Design*. Morgan Kaufmann Series in Interactive Technologies. Morgan Kaufmann Publishers, Amsterdam, Boston, Heidelberg, and others, 2nd edition, 2000.

[37] G. Wesche, J. Wind, M. Göbe, L. Rosenblum, J. Durbin, R. Doyle, D. Tate, R. King, B. Fröhlich, M. Fischer, M. Agrawala, A. Beers, P. Hanrahan, and S. Bryson. Application of the Responsive Workbench. *IEEE Computer Graphics and Applications*, 17(4):10–15, July–Aug. 1997.

[38] D. Wigdor, C. Shen, C. Forlines, and R. Balakrishnan. Perception of Elementary Graphical Elements in Tabletop and Multi-surface Environments. In *Proc. of Human Factors in Computing Systems (CHI)*, pages 473–482, New York, USA, 2007. ACM Press.

[39] B. Yost and C. North. The Perceptual Scalability of Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):837–844, Sept./Oct. 2005.

[40] J. Zhang and D. A. Norman. Rrepresentations in Distributed Cognitive Tasks. *Cognitive Science*, 18(1):87–122, Jan.–Mar. 1994.

[41] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and M. S. T. Carpendale. Heuristics for Information Visualization Evaluation. In *Proc. of the Workshop Beyond Time and Errors (BELIV 2006), held in conjunction with AVI*, pages 55–60, New York, USA, 2006. ACM Press.

An Exploratory Study of Visual Information Analysis

Removed for blind review

ABSTRACT

To design information visualization tools for collaborative use, we need to understand how teams engage with visualizations in their analysis process. We report on an exploratory study of groups of individuals, pairs, and groups of three engaged in visual information analysis tasks using paper-based (static) visualizations. From the analysis of our study, we derive a framework that captures the analysis activities of co-located teams and individuals. We present a comparative analysis of this framework with existing models of of the information analysis process, which suggests that information visualization tools may benefit from providing a flexible temporal flow of analysis actions.

Author Keywords

Information Visualization, Analysis Process, Collaboration

ACM Classification Keywords

[H.5.2]Information Interfaces and Presentation;

INTRODUCTION

Interactive information visualization tools are often the center of many complex information analysis tasks [16]. In everyday practice, data is frequently interpreted and analyzed not only by individuals but by *teams* of individuals working in concert to make decisions. While many researchers have explored the information analysis process (e. g. [3, 5, 12]), little has emerged on the nature of this process in a collaborative context [6, 8]. How a single doctor would analyze biomedical visualizations, for example, might differ from how a team of doctors might analyze the same data. If teams make use of visual information to solve problems differently than individuals, we need to understand what these differences are so we can redesign infovis tools to support their activity. To address this problem, we designed an exploratory study to understand the flow and nature of this collaborative process and its relation to individual analysis practices. To derive practical guidelines for information visualization tool design, we focused on analyzing how participants engage with the workspace and their collaborators. Teams in our study were given paper-based (static) visualizations to solve tasks, allowing us to view their process independently of the confounds of a specific infovis system. The analytic framework that we have derived from our observations allows us to deconstruct and understand this visual information analysis process for the purpose of design, heuristic evaluation, and analysis of information visualization tools.

Our work makes primarily three contributions: first, we present an exploratory study aimed to examine the information analysis process for individuals and small groups in the context of visual data; second, we present an analytic framework that allows researchers to understand this analysis process in other contexts, and finally, we provide three concrete design implications for digital information visualization tools derived from our findings.

BACKGROUND AND RELATED WORK

Since our research study shows individuals and teams solve information tasks using simple visual representations of their data and results in an information processing framework, it particularly relates to previous studies that have also resulted in information processing frameworks. Next, we outline research that articulates an *information visualization process* or the process through which a person extracts insight from a dataset given a problem and visualization tool. Then, we describe other studies that are related in that they also consider team work and more general information handling.

Collaborative Visual Information Processing

Both Park et al.'s study of pairs using distributed CAVE environments [8], and Mark et al.'s study also of pairs but using a shared information visualization software tool [6] have resulted in similar but not identical information processing frameworks. These two studies are most related to ours but our study differs in that by studying non-digital information processing our framework does not reflect the processing constraints built into existing software. A detailed comparison of these frameworks with ours is included in the discussion section.

Without the benefit of associated studies, several other researchers have also modeled a user's involvement in visual information processing as an iterative sequence of components; however, each model is unique in terms of its focus, and how it abstracts the process. These models have focused on *individuals'* use of visualizations—only recently have researchers shifted their focus toward how *teams* use visualizations together.

One perspective has been concerned specifically with the design of digital information visualization tools, focusing on how users manipulate view and visualization transformation parameters, e. g., [4, 5]. Jankun-Kelly et al. propose a model of visual exploration for analyzing a user's interaction with a digital visualization system [5]. The key insight of this work is that *a fundamental operation in the visual exploration process is the manipulation of visualization parameters*. This model is effective in capturing the temporal aspects of visual parameter manipulation; however, it does not capture the higher-level semantics of a user's interaction (i. e., why did the user change that parameter). Chi and Riedl's model [4] addresses these semantics, basing their semantic operator framework on users' intention of action (i. e., view filtering vs. value filtering), classifying and organizing operators in the analysis process.

At the other end of the spectrum, Amar and Stasko name higher-level analytic activities that users of a visualization system would typically perform, such as complex decision-making, learning a domain, identifying the nature of trends, and predicting the future [1]. Shneiderman outlines a two-step process (from overview to detail), that addresses a task-centric perspective on the analysis process. He names seven operations that information visualization tools should support to facilitate the problem solving process: overview, zoom, filter, details-on-demand, relate, history, and extract [11]. Similarly, a model by Russell et al., derived from studying collaborative information consolidation activities, describes a "Learning Loop Complex" [9], a cyclic process of searching for representations and encoding information. Indirectly, these observations have led to Card et al.'s *sense making cycle* [3] (extended in [16]). While models from this latter perspective are more closely related to our processes, most have a stronger cognitive focus. We will later revisit the sense-making cycle by Card et al. [3] as it shared some processes defined in our framework.

In contrast to these two main perspectives we are interested in the *general processes* that occur during analysis (independent of the confines of a computer-based infovis tool), as well as the interactions with visualizations and those between team members during collaborative information analysis. We are interested in general processes that form the basis of *collaborative* information visualization as the low-level mechanics of interacting with an infovis tool are probably not indicative of how teams would solve a visual information problem.

Choosing a Methodology

When developing software tools to augment work practices, at least three fundamentally different approaches exist. One is to study possible improvements for support of the process through studying the current software support or tools in use. Another is to hypothesize about improvements in existing tools, develop a promising tool and study it in comparison to the existing tools. A third is to work towards an improved understanding of the process in order to develop a better match between natural human process and its software support.

Our approach falls into the latter class, and begins with the premise that through observations of users' interactions with physical artefacts, we can develop a richer understanding of basic processes that can be used to inform interface design. Other researchers (e.g. [10, 15]) have taken this approach, studying how groups accomplish tasks in *non-digital contexts* in order to understand what activities *digital tools* should support. The reasoning behind this choice is that users' *physical interactions* with these familiar artefacts and tools would closely reflect how they *understand and think* about the problem at hand. For instance, Tang's study of group design activities around shared tabletop workspaces revealed the importance of gestures and the workspace itself in mediating and coordinating collaborative work [15]. Similarly, Scott et al. studied traditional tabletop gameplay and collaborative design, specifically focusing on the use of tabletop space, and the sharing of items on the table [10]. While these authors studied traditional, physical contexts, ultimately their goal was to understand how to design *digital* tabletop tools. Both of these studies contributed to a better understanding of collaborative work practices involving tables in general.

The approach taken in these two studies works well when addressing a design area where the critical issues are poorly understood. For instance, we are *uncertain* how groups will work together with information visualization if given the ability to do so freely (e.g. prior efforts involved systems where individuals could not work in parallel [8] [6]). Furthermore, we do not know how teams will share and make use of intermediate results, or indeed whether they will even share and work together from the same views or artefacts of the data.

Our work builds on efforts of prior researchers in developing frameworks to understand the visual information analysis process, and the work of researchers attempting to understand collaborative behaviour. The study we describe here takes a tentative first step toward building our understanding of collaborative visual information analysis. We can then leverage this understanding to build infovis tools to support collaboration.

A STUDY OF THE INFORMATION ANALYSIS PROCESS

We conducted an exploratory study to understand the visual analysis process. The study focused on examining how processes differed between individuals and small groups (pairs and groups of three).

Participants

We recruited 24 paid participants from the university population, 14 female, 10 male, primarily from science, arts, social sciences, and business. The mean age of the participants was 26 years. We had 4 groups each of singles, pairs, and triples. With one exception, all pairs and triples were known to each other before hand and did similar data analysis tasks at least as frequently as yearly. For group construction seen Figure 1.

Apparatus

Participants worked on a large table (90 × 150 cm) and were given 15 × 10 cm cards each showing one data chart. The table was covered with a large paper sheet, and several pens,

Scenario	Task	Type
C (Cereal)	1) Give a short description of the participants' characteristics.	open
	2) Who should each breakfast option be advertised to?	open
	3) Do more females prefer oatmeal than active people prefer cereal.	focused
	4) Do more inactive people prefer oatmeal than people over 60? Do you think there might be a relationship between lifestyle and age in terms of preference for oatmeal?	focused
B (Behaviour)	1) Find pairs of behaviours that have similar ratings in at least three different situations.	open
	2) Choose three situations and describe behaviours most appropriate for that situation.	open
	3) Find two situations that have at least five behaviours with similar ratings.	open
	4) Is it more appropriate to argue or belch in a park?	focused
	5) Where was it most appropriate to laugh.	focused
	6) What behaviour in which situation was most appropriate and which was most inappropriate.	focused

Table 1: Study questions and type per scenario.

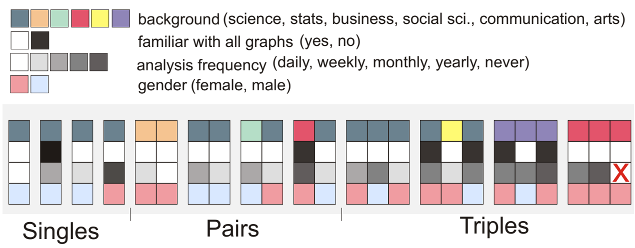


Figure 1: Participants' gender, chart familiarity, and data analysis frequency.

pencils, rulers, erasers, scissors, and sticky notes were provided. Six different types of charts were used. These charts showed different subsets of the data and each data subset was shown in at least two different representations (e. g., line chart and bar chart). Figure 2 gives an overview of the charts used and shows that how many participants reported themselves to be unfamiliar with a given chart; however, even, though some participants were unfamiliar with certain charts no participant reported to be unfamiliar with a chart whose data was not redundantly encoded with another chart familiar to him/her.

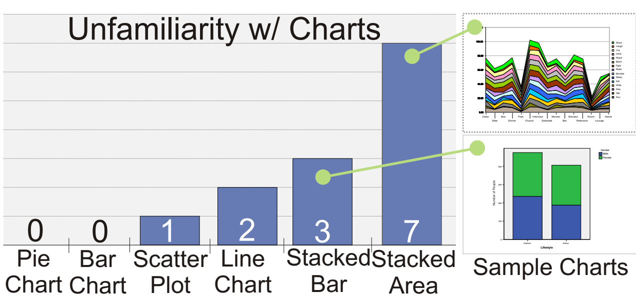


Figure 2: Unfamiliarity of participants with charts in the study.

Tasks

Participants worked on two task scenarios each composed of a different data set with its own representations. The data sets used in the study are part of the sample files provided with the analysis software SPSS 14.0. The behaviour data set (Scenario B, behavior.sav in SPSS) included 32 charts (1 stacked area, 1 line, 15 scatter plots, 15 bar charts). The data shown in these charts was about ratings for the appropriateness of 15 behaviours in 15 different situations (e. g., running in church). The cereal data set (Scenario C, cereal.sav in SPSS) which included 30 charts (3 pie, 9 bar, 9 stacked bar, 9 line charts) was about an imagined study of preferences for certain breakfast options. The presentation order of these scenarios was counter balanced between groups. Similar to the design used in [6], our scenarios each contained an equal number of open discovery tasks, where tasks could have several possible solutions, and focused question tasks which had only one correct answer. Table 1 gives an overview of the study tasks.

Procedure

Participants were greeted and then seated themselves around the table. Next a short tutorial was provided on the types of charts, tasks, and scenarios used in the study. Participants were told that they could use any of the tools (pens, rulers, etc.) to work with the charts, and that they could write on anything as they saw fit (e. g., cards, scrap paper, table, etc.). Participants were then given an example task scenario to clarify the process. Once it was clear how to proceed, each task scenario was given in turn, and the participants were instructed to work on the tasks in any way they felt comfortable. Upon completing both task scenarios, participants filled out a questionnaire asking about their experiences during the study and to collect demographic information. The groups of two and three participants naturally discussed their tasks and progress and single participants were asked to use a talk aloud protocol.

Data Collection and Analysis

During each session two observers were always present. Both observers collected notes, and each session was video or audio taped. 610 minutes of video data was collected (≈50 minutes for each session). Our multi-pass open coding analysis was based on both the collected notes and the video data. Notes were used to form initial coding categories which were used for the first video coding pass and were refined through subsequent study of the videos and the second coding pass. This process, similar to that used in [15, 10], provides a rich understanding of the information processing and analytic activities within groups, of the similarities and difference across groups and of the unique character of each group.

FINDINGS

In this section, we outline our understanding of the collaborative and individual visual analysis process as revealed during our analysis. We follow this by illustrating how the processes themselves were *not* temporally organized in a consistent way across groups, suggesting that information visualization tools should consider supporting this flexibility. In the next section, we relate these findings to prior work, and discuss how they can inform the design of information visualization tools.

Processes in Visual Information Analysis

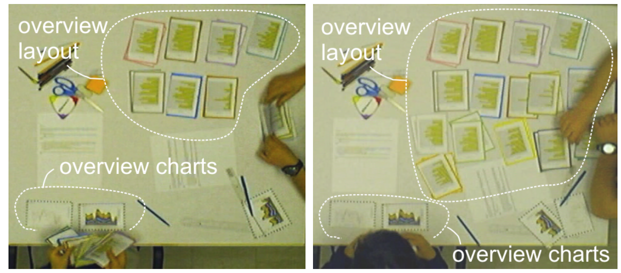
Our analysis revealed eight processes common to how participants completed the tasks in our study: browse, parse, discuss collaboration style, establish task strategy, clarify data, operate on data, select data, and validate findings (summarized in Table 2). We describe each process using examples drawn from our study, discussing participants' interactions with one another and the workspace and elaborate on how the processes differed between group types. Where average process times are reported these need to be read as an accumulation of several instances of particular processes during both scenarios.

Browse:

The browsing process comprised activities involving *scanning through data to get a feel for the available information*. Browsing activities did not involve a specific search-related to a task; instead, the main goal appeared to be to gain some understanding of the data set. For example, we observed participants quickly glancing through or scanning the information artefacts—likely to see what types of charts were available and the variables in the charts. Five participants took the complete pile of charts and flipped through them in their hands, while eleven others created an elaborate layout of cards on the table. Figure 3 shows an example in which two participants use two very different browsing strategies. One participant (bottom of image) lays the two overview charts out in front of him, flipping through the remaining cards in his hand, while the other participant creates a small-multiples overview of the cards on the table as he browses through them one at a time. Groups were slightly more efficient than individuals (average browsing times were $\approx$30s for groups, and $\approx$60s for individuals), perhaps indicating for individuals, having a completely clear sense of the data is more important, whereas groups can rely on others. In one case, we observed one participant in a group of three who did not "browse" through the data himself; instead, he watched as his partners laid their cards out on the table.

Parse:

The parsing process captures *the reading or re-reading of the task description* in an attempt to understand how to solve the problem. Participants read the task description both quietly or aloud, and in teams, this choice reflected the collaboration style that teams adopted: for instance, teams working closely together would read task descriptions aloud, facilitating joint awareness of the state of the activity, and discussion of how to interpret the question. These readings would sometimes result in a rephrasing of the question or note-taking of required variables. On average, groups spent 2.5 min reading and re-reading the task description, regardless of the size of the group; however, individuals referred to the task sheet



(a) Start of a browsing session. (b) End of a browsing session.

Figure 3: Different browsing strategies: the participant on the right creates an overview layout; the participant on the bottom laid out the overview charts and is flipping through the remaining data charts in his hands.

more frequently (10 times vs. 9 times for pairs and 7 times for triples).

While many real-world information analysis scenarios may not have a concrete problem description sheet, an assessment of the given problem(s) and the required variables can certainly still occur and would be considered part of this process. The problem sheet can be seen as external textual information that is not part of the current dataset but provides meta information on the problem, tasks, or data.

Discuss Collaboration Style:

Many teams explicitly discussed their *overall task division strategy*. We observed several collaboration strategies ranging from completely independent to closely coupled work:

- *Complete task division*. Participants divided tasks between themselves so that they would not duplicate work. Each participant worked alone with his or her information artefacts on a pre-specified subset of the problems. Results would then be combined at the end without much further group validation.
- *Independent, parallel work*. Participants worked on each task independently, but at the same time. When one participant had found an answer, solution and approach were compared and discussed with the group. Other participants might then validate the solution by retracing the approach with their own artefacts, or by carefully examining the partner's information artefacts.
- *Joint work*. Participants talked early about strategies on how to solve the task, and then participants went on to work closely together (conversing and providing assistance) using primarily their own or shared information artefacts. When one person found a solution, information artefacts were shared and solutions were validated together.

Interestingly, while teams might explicitly discuss a collaboration style, *all eight teams* changed their collaboration strategy midway through a task scenario or between scenarios. A combination of parallel and joint work strategies was used by six teams and two others used a combination of task division/parallel and task division/joint work. Six of the eight teams started with a loose definition of doing the tasks "together." Strategy discussions were typically very short: $\approx$2 min on average per scenario. Most of the changes in task

Process	Description	Goal
Browse	scanning through the data	get a feel for the available information
Parse	reading and interpretation of the task description	determine required variables for the task
Discuss Collaboration Style	discuss task division strategy	determine how to solve the tasks as a team
Establish Task Strategy	establish how to solve a task with given data & tools	find an efficient way to solve the problem
Clarify	understand a visualization	avoid mis-interpretation of the data
Select	pick out visualizations relevant to a particular task	minimize the number of visualizations to read
Operate	higher-level cognitive work on specific data view	solve task or sub-task
Validate	confirm a partial or complete solution to a task	avoid errors in completing the task

Table 2: The eight processes in information analysis. "Discuss Collaboration Style" only applies to collaborative analysis scenarios.

strategy were quite seamless, and did not require any formal re-negotiation. This is echoed in the post-session questionnaire in which two participants reported to have chosen their strategy "intuitively" and "by chance." In general, teams showed a strong tendency for parallel work: all eight groups solved at least parts of one scenario in parallel. 14 of 15 participants reported that the main reason they divided tasks this way was for perceived efficiency.

Establish Task Strategy:

In this process, participants *searched for the best way to solve a specific task using the given data and tools*. The goal of establishing such a strategy was to determine the next views or interactions required to extract variables or patterns from the data to solve the problem efficiently. As a team activity, this discussion occurred often with the help of individual information artefacts. On many occasions, one participant would present a possible approach to the other participant(s) using examples. For example, Figure 4 illustrates an instance where two participants are discussing how to solve a particular task using a specific chart they had chosen. The team frequently flipped between looking at a shared chart and the chart in their own hand. This explicit strategy discussion was more common when teams worked in a *joint work* collaboration style. When participants worked independently or in parallel, the determination of strategy seemed to occur silently (perhaps in parallel to the *parsing* process). For instance, participants might articulate their strategies without discussing the explicit reasoning for it: "I am now going to look for the highest peak." During the video analysis, we only observed on average 1-2 minutes per scenario in which teams specifically discussed their strategy to solve a task.

At the end of this process, depending on the chosen strategy, participants often reorganized their information artefacts in the space to create an adequate starting position for solving the task. For example, if the strategy was to find two data charts, then the workspace might be organized to facilitate the finding of these two data charts (as in Figure 3).

Clarify:

Clarification activities involve efforts to *understand an information artefact*. While we provided users common bar, pie, and line charts, we also provided less commonly used stacked bar charts and an area chart. The unfamiliar charts required more careful scrutiny by participants. For individual participants, ambiguities in the data display were often resolved using other charts as aids, by re-reading parts of the scenario or task descriptions, through annotations on information artefacts, and in one case, the drawing of example diagrams. In teams, the need for clarification additionally involved discussion with other participants to decipher and understand the charts and sharing of information artefacts. Overall clarification required less than 1min for Scenario B and no clarification was required for Scenario C. The clarification times for Scenario B were higher for all groups as this scenario contained the most unfamiliar stacked area chart. Only those triples that included participants which were unfamiliar with certain charts required longer than average (1min, 2min) for clarification in Scenario B.

Figure 4: Discussing a strategy on how to solve a task using the chosen chart. Information artefacts are used as aids.

Select:

Selection activities involved *finding and picking out information artefacts relevant to a particular task*. We observed several different forms of *selection*, often dependent on the organization of data that was established during *browsing*. We characterized these styles of selection by how artefacts were *spatially separated* from one another:

- *Selection from an overview layout*. Beginning with an overview layout (e. g., small-multiples overview from Figure 3), relevant cards are picked out. Selection of cards from this layout involved either a *re-arrangement* of the organization scheme so that relevant cards were placed within close proximity or *marking* by either placing hands or fingers on the cards, or using pens.
- *Selection from a categorization layout*. Beginning from a pile-based categorization of information artefacts, piles are scanned and relevant cards are picked out. These cards are then placed in new piles that carry semantic meaning (e. g., relevant, irrelevant, ...). Previously existing piles might change their meaning, location, and structure in the process.

How users organized these selected data cards was dependent on how they intended to *operate on* (or use) them. The left of Figure 5 illustrates an instance where two cards to be compared were *relocated* and placed side-by-side. The right of Figure 5 shows an example where a variable was to be measured, so the card was *relocated* closer in the individual person's workspace. Frequently, the spatial organization of cards relative to piles of data in the workspace carried semantic meaning. For example, when an *operation* on a data card was to be brief, a single card was drawn out, operated upon, and then replaced. Similarly, the organization scheme might reflect the perceived importance of a set of cards: at times, we observed piles of information artefacts that were clearly discarded (Figure 6). Temporally, we also observed different *selection* strategies, which could be loosely classified as "depth-first" or "breadth-first." A "depth-first" approach involved selecting a single card, *operating* on it for a period of time, and then selecting the next card (e. g., Figure 6, left). "Breadth-first" strategies selected all cards deemed relevant in a single pass and then *operated* on them afterwards (see Figure 6, right). On average participants spent $\approx 4min$ selecting data, the second most common process in our study.

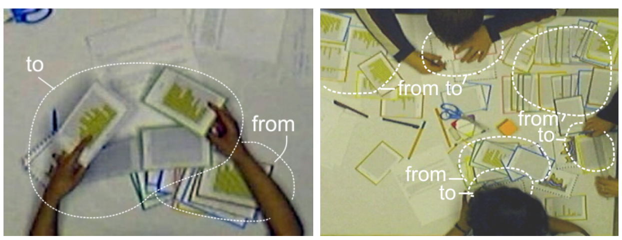


Figure 5: Chart organization during selection depending on their intended usage. *Left*: a participant selected four cards for comparison placing them side by side in her hand. *Right*: three participants selected individual charts and placed them in the center of their workspace to measure a specific value.

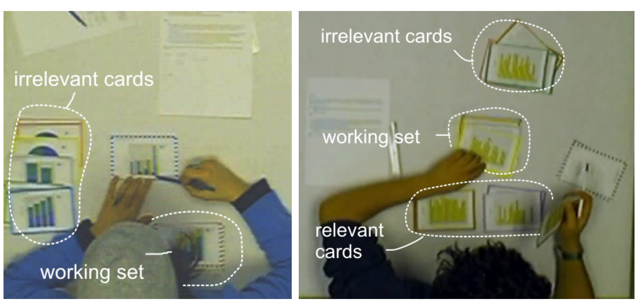


Figure 6: Changing categorization during selection. *Left*: a participant placed irrelevant cards to her left and picks single cards to operate on from the working set. *Right*: a participant picked out relevant cards, placed them close to himself, and put irrelevant cards in a pile further away.

Operate:

Operation activities involved *higher-level cognitive work on a specific view of the data* with the goal of extracting information from the view to solve the task. Figure 7 illustrates the two most common types of operation activities: extracting a data value, and comparing data values. To extract a data value from a card, participants often used rulers or some other form of measuring tool (e. g., edge of a piece of paper). To aid recall of these values, participants often made annotations: sometimes on the charts themselves, and other times on spare pieces of papers. During the course of both scenarios each participant on average annotated at least three information artefacts (2 during Scenario B, 1 during Scenario C). Every participant in our study compared charts on at least one occasion. The most frequent comparison involved just two charts but we also noted 15 occasions of participants comparing three or more charts. In our study, participants arranged the charts for a comparison during *selection*: cards would be placed in close proximity to facilitate easier reading of either individual values or patterns (Figure 6). Participants were quite creative in their use of tools to aid comparison: marking individual values, bending or cutting individual charts (to facilitate placing values physically side-by-side), or on 7 occasions we noted overlaying of charts atop one another in an attempt to see through the top chart. The operation process typically generated a set of results, which were synthesized with previous results and/or written down. During team activity, results were sometimes reported to the team if other tasks depended on these results (e. g., during joint activity). Operation was the most time-consuming activity in our study. On average participants spent almost half of their time (11 minutes) on operations per scenario. For all groups and scenarios operations most frequently followed a selection process.

Figure 7: Two participants showing two different types of operations on the information. The participant on the right is comparing two cards using a ruler while the participant on the top is measuring a particular value.

Validate:

Validation activities involved *confirming a partial or complete solution to a task*. Beyond confirming the correctness of a solution, teams also ensured the correctness of the process or approach that was taken. In teams, the validation process often included discussion coupled with sharing of information artefacts: some participants validated others' solutions by looking carefully at the solution (in terms of the information artefacts), while others validated the solution by using their own information artefacts (i. e., the process or approach was shared instead of the artefacts themselves). When working more independently, the validation process only involved the presentation of a solution by the group member who had found the solution. In groups where collaborators worked more closely, the collaborators would often ensure that the other participants had understood the process with which a solution was found. For individual participants, the validation process involved looking at other data cards (i. e., different representations) for the same answer. Of interest is that individuals appear to be concerned about the "correctness" of their solution/approach based on other information artefacts, while teams also rely on a collective validation from the social group. On average groups of three spent the longest time validating their answers (≈ 3min), pairs spent ≈ 2min validating, and individuals spent less than one minute validating their answers.

Temporal "Sequence" of Processes

To understand how the processes related to one another in terms of a temporal relationship, we analyzed the video data from our study, coding each individual's activities using these process labels. This analysis revealed three aspects of participants' activity: first, while certain processes frequently occurred before others (e. g., *select* most frequently appeared before *operate*), *no common overall pattern appeared*; second, *individuals* varied in how they approached each task, and finally, teams also varied drastically in how they spent their time. For brevity, we present a few example charts. All charts for singles, pairs, and triples exhibit this same extreme variability of approach.

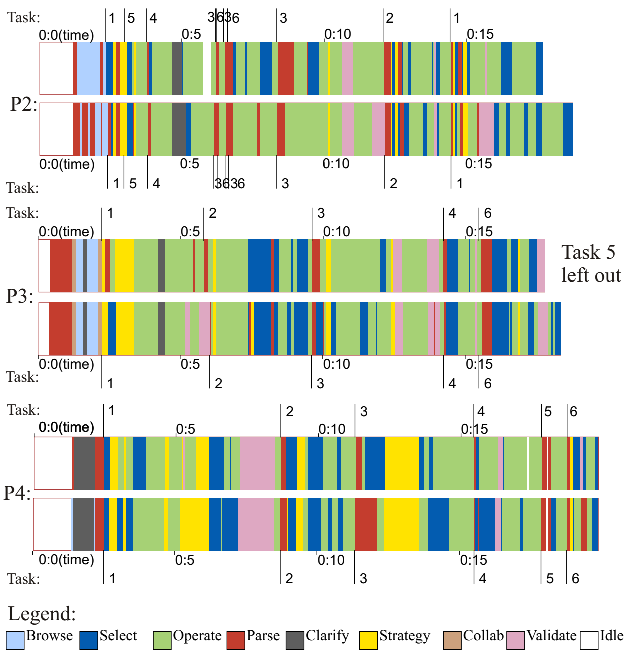


Figure 8: Temporal sequence of processes for three pairs during one complete scenario.

Figure 8 shows the coded temporal sequence of analytic processes during Scenario B for three pairs. Notice how the sequence of processes was quite different for each pair, even though participants worked on the same tasks using the same tools, representations, and views of the data. Even within teams participants did not show the same temporal occurrences of processes. On average participants in pairs were concurrently working in the same process for ≈ 70% of the time. For Scenario B (Figure 8), P2 has a 65% co-occurrence of the same processes, P3 80%, and P4 69%. This reflects the collaboration strategies participants had chosen. P3 had switched from a complete task division to joint work in this scenario while P2 and P4 were working mostly in parallel. Participants in groups of three only showed a 40% co-occurrence of processes on average.

In Figure 8, Tasks 1–3 were open discovery tasks and Tasks 4–6 were focused question tasks. We noticed that both individuals and teams solved focused question problems quicker than open discovery tasks. Teams had a better understanding of the tasks (established during the task strategy process) and solved them (both focused and open discovery tasks) more correctly. This result echoes findings in [6] that suggest that groups perform more accurately, albeit slower. Of course, teams also exhibit *establishing a task strategy* more so than individuals, again in order to establish common ground, or to ensure a correct or agreed-upon approach.

Figure 9 shows a detail view of a specific task, charting individual participants and three of the participant pairs. Notice that even for a single task occurring over a roughly five minute sequence, *how* the participants engaged in the task, and the temporal distribution of process time varied.

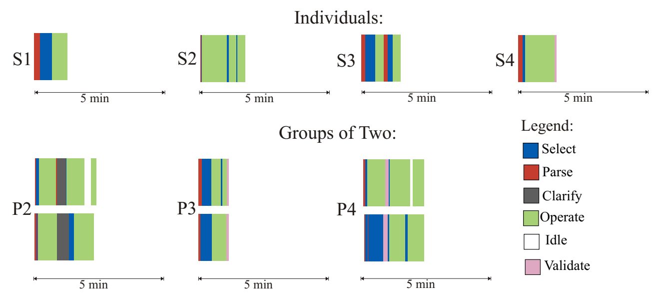


Figure 9: Temporal sequence of processes for one open discovery task. The top row shows timelines for individual participants (S1–S4). The bottom row holds timelines for participants in groups of two (P2–P4).

DISCUSSION

To this point, we have introduced a set of processes that occur within the context of collaborative and individual visual information analysis. These processes apparent from our study form an eight-process framework. The framework is unique from prior work in that it provides an understanding of how teams and individuals use information artefacts in the workspace to solve visual information analysis tasks and of how team members engage with each other during this process. In this section, we discuss how our framework relates to other information analysis/information visualization models. This discussion reveals that while individual processes relate closely to existing models, our temporal analysis suggests that with appropriate tools, both the collaborative and individual information analysis processes may naturally be more fluid and benefit from temporal flexibility.

Comparing Frameworks

Comparison with the Sense-Making Cycle

Card et al. [3, pp. 10] provide a high-level model of human activity called the "knowledge crystallization" or "sense-making cycle" where the goal is to gain insights from data relative to some task. This model includes five main components: **foraging for data, searching for a schema** (or representational framework), **instantiating a schema, problem solving, and authoring, deciding or acting**. It builds on work by Russell et al.[9] which involved observations of collaborative work and an extension can be found in [16]. Spence [12] extends this model by specifically exploring the "foraging for data" component in terms of visual navigation. In particular, he relates visual navigation to *cognitive activities* (such as internal model formation and information interpretation), thereby arguing that how users can navigate, explore, and visualize a data space will shape how users think about the data.

The Sense-Making has several components related to our model. It outlines a process called "foraging for data" that includes our *browse* process. Spence distinguishes three different browsing activities [13]: *exploratory browsing* where the goal is to accumulate an internal model of part of the viewable scene; *opportunistic browsing* to see what is there rather than to model what is seen; and *involuntary browsing* which is undirected or unconscious. We primarily observed exploratory browsing, and saw that as part of this process, participants established a layout of cards, or put cards in observable categories (e. g., by variables or graph types). It seemed that those participants that created a specific layout of cards in their work area created a type of overview by imposing an organization (even if a loose one) on the information artefacts. Thus, we saw a physical manifestation of the creation of an "internal model of the data." Furthermore, these physical layouts (a consequence of the browsing phase) clearly relate to Shneiderman's "overview" task [11].

"Search for schema" seems to involve activities that we characterize as being a part of *parsing*, specifically the identification of attributes on which to operate later. The activity of identifying attributes to look for in the data described in this model is augmented in our parse component by additional activities of discussion, and note taking found during our study.

In Card et al.'s model "search for a schema" and "instantiate schema" involve activities that help in the search for the best way to solve the given problem with the provided visualization tool and therefore relate to our *task strategy* process, albeit being more tool-centered than our definition.

Clarification is not an explicit component in this model but the need for clarification would typically arise during the searching for and instantiating a schema components. Our *selection* process is most closely related to the foraging for data component but can extend into the searching for and instantiating a schema components when participants have ended their browsing activities and are ready to select specific information important to solving the task. This models include activities that we see as part of an operation process: problem-solving, including Bertin's three levels of reading: read fact, read compare, read pattern [2]. *Validation* is not directly represented in Card et al.'s model [3]; perhaps, as we have also observed, because validation seemed to be often omitted or quite brief for individual participants and their model focuses on a single user.

The sense-making cycle is the most highly coupled and interactive of the three models we are comparing to. However, it makes a strong temporal (cyclical) suggestion but does allow for loops within this cycle over defined forward and backward connections between individual components. In general, the sense-making cycle is not identical to our model but predicts some of our findings in terms of temporal flexibility and shares some components with our model.

Collaborative Analysis Models

In studying pairs using distributed CAVE environments, Park et al. articulate a five-stage pattern of behaviour: **problem interpretation, agreement on vis tool to use, search for a trend, discovery reporting, and negotiation of discoveries** [8]. Mark et al. also provide a five-stage collaborative information visualization model: **parse question, map 1 variable to program, finding correct visualization, validating the visualization, and validation of the entire answer** [6]. A loop is included for additional variables from stages four back to stage two. The temporal sequence of stages in this model was derived from a study of pairs solving both free data discovery and focused question tasks in both distributed and co-located settings. These two models share some similarities, but are clearly not identical. A possible explanation for the disparity is that Mark et al.'s model [6] focuses on a context where the pair negotiates exploration through a *shared* tool (i. e., they could not work in a decoupled fashion [14]) whereas Park et al.'s model [8] allows for more loosely coupled work.

Both models share some similarity in the processes discovered in our study. Our *parsing* process relates closely to Mark et al.'s *parse question* [6] and Park et al.'s *problem interpretation* [8] stages. We augment these stages with activities that might not have been part of the specific environment under study in both models: note taking and frequent discussion about how to interpret a certain task. The *discussion of the collaboration style* is not explicitly covered in both models. However, similar to Park et al.'s study we observed a strong tendency in all group conditions for participants to do at least part of the work using their own views and information artefacts. Similar differences in work styles for spatially fixed information visualization tasks (e. g., maps that cover the whole workspace) have been described in [14], but they have not been put in a greater context of other processes of visual analysis. According to Mark et al.'s model, "map 1 variable to program" is closely related to our *task strategy* process in that it would also involve a collaborative agreement on the most appropriate visualizations, parameters, or views to solve the problem [6], like Park et al.'s *agreement on visualization tools to use* [8]. However, our description of this process discusses the activities involved in establishing a strategy rather than describing it in the context of a specific tool.

In contexts where new visualizations are introduced, or individuals are brought in without prior training on particular visualizations, the need for *clarification* would be common. Specifically, beyond providing users with aid in developing an understanding of a particular visualization, we would expect individuals to ask for collaborators' interpretations of that visualization or interaction technique or to put their own views and interpretations up for discussion. Considering clarification as a process of analysis is important for designing and evaluating visualization tools but it is not a specific part of the two collaborative analysis models.

Our articulation of the *selection* process is related to parts of the activities covered by Mark et al.'s "find correct visualization" stage and Park et al.'s "search for trend." Our description of *selection*, however, more broadly captures the notion of picking out important information beyond operations in a specific visualization system.

"Independent search for a trend including some adjustments to viewing parameters" and "report discovery" include *operations* as defined in our model. Operation is not an individual stage in Mark et al.'s model but is integrated in the "find correct visualization" stage [6]. In groups, the validation stage was much more visible and it is also included in these two models as the last stage of information analysis [6, 8]. Mark et al. noticed differences in validation between the free discovery and focused question tasks; a result that was echoed in our study. During more open-ended questions, validation was usually longer and involved more discussion than for focused tasks.

In general, both these models are related to ours in that they share some of the processes discovered in our study but are quite different in their suggestion of a fixed temporal order.

Temporality and Process-Free Tools

Many of the existing models suggest a *typical* temporal order of components; however, our analysis of the temporal occurrence of the framework processes in our study suggests that this typical temporal ordering was not evident. We argue that our finding of a lack of a common temporal ordering reflects the design of our study; in particular, the stipulation that participants would use a paper-based "information visualization" tool along with traditional tools such as pens, paper and notepaper. Traditional tools have no specific flow in terms of which tools should be used first or for what purpose (in contrast, typical interactive information visualization tools require specific ordering of interactions to get specific visualization results). As a consequence, we argue that the processes and interactions we observed with these traditional tools better reflect the thought and collaborative processes. We believe that prior authors' finding of a common temporal ordering more likely reflects the use of information visualization tools with a specific *process-flow*.

The flexibility afforded by traditional tools allowed individuals to approach tasks differently. As a consequence, they also allowed groups to transition between multiple stages of independent and closely coupled work rather than regimenting particular work process.

In summary then, the processes in our analytic framework map to related models, yet our analysis suggests that the temporal ordering of these components is by no means universal. In many digital information visualization systems, the flow of interaction is regimented by structure; in contrast, the use of traditional tools in our study allowed participants to freely choose how to approach and solve problems. On this basis, we believe this analytic framework can be used as a means to understand information visualization tools: for example, to asses temporal or procedural work processes that a particular system might impose.

IMPLICATIONS FOR DESIGN

Most information visualization systems have been designed for a single user, but co-located collaborative analysis of information is also common. Until relatively recently people have had to rely on physical prints of information for co-located collaborative analysis. The emergence of large, interactive displays opens new possibilities for the development of interfaces to support collaborative analysis using information visualizations. In this section, we discuss implications for the design of single-user and co-located multi-user information visualization systems based on our findings.

Support Flexible Temporal Sequence of Work Processes:

Individuals have unique information analysis practices based on their prior experiences, successes, and failures. These well-established work practices should be supported by digital systems. Our study showed that all participants worked differently in terms of the order and length of individual work processes they engaged in, suggesting the need for digital systems to be relatively unrestricting. The temporality of work processes suggested by previous models of the analytic process could imply that common information visualization tools require a specific process-flow. Our study, however, suggests that users of digital systems may benefit if a flexible order of operations can be performed. Co-located collaborative systems, in which more than one user may work and interact at the same time, should possibly *allow group members to be engaged in different types of processes at the same time* and also allow them to work together adopting the same processes.

Support Changing Work Strategies:

In group settings, our participants *dynamically* switched between closely coupled and more independent work. The *browse, parse, operate,* and *select processes* were most often done on individual views of the data in a more loosely coupled fashion. *Discussion of collaboration style* and *establish task-specific strategy, clarify,* and *verify* often happened in closer cooperation with the other partner(s) and often included shared views of the data. To support these changing work strategies information visualization tools for co-located work need to be designed to *support individual and shared views of and interactions on the data*. Each collaborator should be able to perform individual operations on these views unaffected by his or her team members' actions. However, the tool should also help to share these individual views and, thus, provide awareness of one team member's actions to the other collaborators. To support individual views of the data, interaction with the underlying data structures (deletion of nodes in a tree, change of query parameters, etc.) should be designed so as to not influence others' views of the same data. However, to support shared views of the data, these previous operations should be transferable to group views, for example, to combine highlights, annotations, or other parts of an interaction history.

Support Flexible Workspace Organization:
The organization of information artefacts on the table changed quite drastically for most of our participants. We observed that participants had quite distinct individual workspaces on the table in which they laid out their cards. These workspaces were quite flexible and would change depending on tasks as well as, in group settings, on team members' spatial needs. This observation is echoed by the studies of collaborative behavior reported in [10] that call for co-located collaborative systems to *provide appropriate functionality in these personal workspaces* (territories). We refer to their paper for further guidelines of how to support personal territories for co-located collaborative work.

Participants also seemed to frequently impose categorizations on data items by organizing them spatially in their workspaces. During *browsing*, overview layouts were often created in which the cards were spread over the whole workspace. Mainly during *selection* and at the end of an *operation* process, information artefacts were organized in piles in the workspace. These piles seemed to have inherent categories and varied greatly in size, lifespan, and semantic. Allowing users to *impose a spatial organization of the information artefacts* in the workspace should be considered in the design of information visualization systems. These spatial organizations can help users support their mental model of the available information. Systems like CoMotion [7] are already taking a step in this direction but the typical information visualization system still relies on a fixed set of windows and controls that can rarely be changed, piled, or relocated.

CONCLUSION

Several researchers have contributed to creating a theoretical understanding of how individuals make use of information visualizations to gain insight into data and solve problems. In this paper, we have continued our evolving theoretical understanding of this process by presenting a framework for visual information analysis. Our framework is based on findings from an observational study that was designed to uncover the processes involved in *collaborative* and *individual* activities around information visualizations in a non-digital setting. We identified eight processes as part of this framework: *Browse, Parse, Discuss Collaboration Style, Establish Task-Specific Strategy, Clarify, Select, Operate, and Validate* and described differences in team and individual work during these processes. We have shown how these eight processes relate to other models of information analysis, and provided insights on differences and commonalities between them. Yet, while others have posited a general temporal flow of information analysis, our results suggest this temporal flow may simply reflect an assumption in the design of existing information visualization tools. Thus, we argue that designers should allow for individuals' unique approaches toward analysis, and support a more flexible temporal flow of activity. These eight processes can, therefore, be seen as an analytic framework that has implications for the design, heuristic evaluation, and analysis of individual and collaborative information visualization systems. In summary, we have furthered the theoretical understanding of information analysis processes, provided a framework to be considered in the evaluation and design of collaborative information systems, and given concrete design implications for digital information visualization systems derived from our findings.

REFERENCES

1. R. A. Amar and J. T. Stasko. Knowledge Precepts for Design and Evaluation of Information Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):432–442, July/August 2005.
2. J. Bertin. *Semiology of Graphics: Diagrams Networks Maps*. The University of Wisconsin Press, Madison, WI, USA, 1983. Translation of: Sémiologie graphique, 1918.
3. S. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings In Information Visualization: Using Vision To Think*. Morgan Kauffman Publishers, Inc., San Francisco, USA, 1999.
4. E. H.-H. Chi and J. T. Riedl. An Operator Interaction Framework for Visualization Systems. In *Proc. of the Symposium on Information Visualization (InfoVis)*, pages 63–70, Los Alamitos, USA, 1998. IEEE Comp. Society.
5. T. Jankun-Kelly, K.-L. Ma, and M. Gertz. A Model and Framework for Visualization Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):357–369, March/April 2007.
6. G. Mark and A. Kobsa. The Effects of Collaboration and System Transparency on CIVE Usage: An Empirical Study and Model. *Presence*, 14(1):60–80, February 2005.
7. MayaViz. CoMotion Discovery. Website, Accessed April 2006. http://www.mayaviz.com/.
8. K. S. Park, A. Kapoor, and J. Leigh. Lessons Learned from Employing Multiple Perspectives In a Collaborative Virtual Environment for Visualizing Scientific Data. In *Proc. of the Conference on Collaborative Virtual Environments (CVE)*, pages 73–82, New York, USA, 2000. ACM Press.
9. D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The Cost Structure of Sensemaking. In *Proc. of the Conference on Human Factors in Computing Systems (CHI)*, pages 269–276, New York, NY, USA, 1993. ACM Press.
10. S. D. Scott, M. S. T. Carpendale, and K. M. Inkpen. Territoriality in Collaborative Tabletop Workspaces. In *Proc. of the Conference on Computer-Supported Cooperative Work (CSCW)*, pages 294–303, New York, USA, 2004. ACM Press.
11. B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proc. of the Symp. on Visual Languages*, pages 336–343, Los Alamitos, USA, 1996. IEEE Comp. Society.
12. R. Spence. A Framework for Navigation. *International Journal of Human-Computer Studies*, 51(5):919–945, November 1999.
13. R. Spence. *Information Visualization*. Pearson Education Limited, Harlow, England, 2nd edition, 2007.
14. A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative Coupling over Tabletop Displays. In *Proc. of the Conference on Human Factors in Computing Systems (CHI)*, pages 1181–1290, New York, USA, 2006. ACM Press.
15. J. C. Tang. Findings from Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies*, 34(2):143–160, February 1991.
16. J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, August 2005.