

An Exploratory Study of Visual Information Analysis

Petra Isenberg
Dept. of Computer Science
University of Calgary
Calgary, AB, Canada
petra.isenberg@ucalgary.ca

Anthony Tang
HCT Lab
University of British Columbia
Vancouver, BC, Canada
tonyt@ece.ubc.ca

Sheelagh Carpendale
Dept. of Computer Science
University of Calgary
Calgary, AB, Canada
sheelagh@ucalgary.ca

ABSTRACT

To design information visualization tools for collaborative use, we need to understand how teams engage with visualizations during their information analysis process. We report on an exploratory study of individuals, pairs, and triples engaged in information analysis tasks using paper-based visualizations. From our study results, we derive a framework that captures the analysis activities of co-located teams and individuals. Comparing this framework with existing models of the information analysis process 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 (infovis) tools are often the center of many complex information analysis tasks [20]. In everyday practice, data is frequently interpreted and analyzed not only by individuals but by *teams* of individuals working in concert to make decisions. Imagine a team of geologists looking at test results to plan an upcoming expedition, a group of city planners examining census data and its influence on future development, or a team of businessmen looking at current data and forecasts of their industry sector. While many researchers have explored the information analysis process (e. g. [3, 9, 16]), little has emerged on the nature of this process in a collaborative context [10, 12]. 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 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 infovis tools.

Our work makes three primary contributions: we present an exploratory study 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 infovis tools derived from our findings.

BACKGROUND AND RELATED WORK

Collaborative Information Visualization Tools

Most collaborative information visualization systems have been developed for distributed data analysis: Many Eyes [21] and Swivel (<http://www.swivel.com/>) are two systems that are targeted at an internet-scale audience and both support *asynchronous distributed* collaborative sharing and exploration of data by letting users upload data, create visualizations, and comment on available visualizations. The Co-Motion environment has been used for information analysis and decision-support applications using shared views of the data on which all users can interact *synchronously* from remote desktops [11]. A collaborative tree comparison system and a set of design guidelines for *co-located* data analysis has been presented in [8]. In relation to this work, our focus is to gain an understanding of the processes of *co-located* collaboration around information visualizations using a single shared workspace. This work particularly relates to previous studies that have also resulted in information processing frameworks as outlined in the next sections.

Collaborative Visual Information Processing

Both Park et al.'s study of pairs using distributed CAVE environments [12], and Mark et al.'s study of pairs sharing an infovis software tool that had been designed for single person use [10], resulted in similar but not identical information

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2008, April 5–10, 2008, Florence, Italy.

Copyright 2008 ACM 978-1-60558-011-1/08/04...\$5.00.

processing frameworks. These two studies are most related, but our results differ 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.

Several researchers have modeled an individual 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. One perspective has been concerned specifically with the design of digital information visualization tools, focusing on how a person manipulates view and visualization transformation parameters, e.g., [5, 9]. Jankun-Kelly et al. [9] propose a model of visual exploration for analyzing one person's interaction with a digital visualization system. A core proposition 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 person's interaction (i.e., why was a parameter changed?). Chi and Riedl [5] address this aspect, basing their semantic operator framework on a person's 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 [1] name higher-level analytic activities that a person using a visualization system would typically perform, such as complex decision-making, learning a domain, identifying the nature of trends, and predicting the future. Shneiderman [15] outlines a two-step process ("overview then detail"), that addresses a task-centric perspective on the analysis process. He suggests seven different operations that information visualization tools should support to facilitate the problem solving process: overview, zoom, filter, details-on-demand, relate, history, and extract.

A model by Russell et al. [13], derived from studying collaborative information consolidation activities, describes a "Learning Loop Complex," 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 [20]). A number of research directions in other domains have asked questions about how collaborators share and coordinate their efforts to work together (see [4]). While these models relate to our research question, most have a stronger cognitive focus. We will later revisit the sense-making cycle by Card et al. [3] as it shares some processes defined in our framework.

Instead of focusing on either task driven or meta-cognitive processes, we are interested in the *general processes* that occur during collaborative information analysis (independent of the confines of a computer-based infovis tool), as well as the interactions with visualizations and those between team members. 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 to existing tools, to 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 the natural human process and its software support.

Our approach falls into the third class. It begins with the premise that through observations of people's interactions with physical artefacts, we can develop a richer understanding of basic work processes that can be used to inform interface design. Other researchers (e.g. [14, 19]) 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 people's *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 [19] revealed the importance of gestures and the workspace itself in mediating and coordinating collaborative work. Similarly, Scott et al. [14] studied traditional tabletop gameplay and collaborative design, focusing on the use of tabletop space and the sharing of items on the table. While these authors studied traditional physical contexts, ultimately their goal was to understand how to design *digital* tabletop tools. Both 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 visualizations if given the ability to do so freely (e.g. prior efforts involved systems where individuals could not work in parallel [12, 10]). 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 prior efforts 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 first step toward building our understanding of collaborative visual information analysis. We can then leverage this understanding to build infovis tools that 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 individuals and small groups in this process.

Participants

We recruited 24 paid participants from the university population, 14 female, 10 male. 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

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

Table 1. Study questions and type per scenario.

to each other beforehand. For group details refer to Figure 1. Our sample size was informed by emerging results. After four pilot studies and 12 groups we were confident that further observations would result in redundant data.

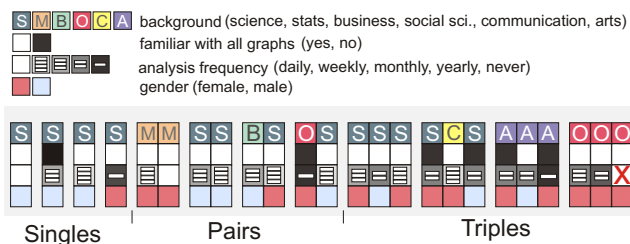


Figure 1. Participants' gender, chart familiarity, and data analysis frequency.

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, 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 how many participants reported themselves to be unfamiliar with a given chart; however, data was always redundantly encoded in familiar charts.

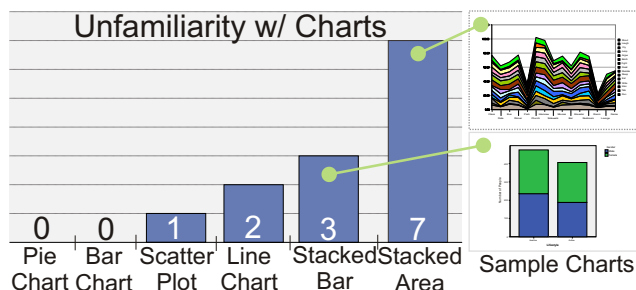


Figure 2. Unfamiliarity of participants with study charts.

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. No specialized knowledge about the data was required to solve the tasks and high task engagement was evident throughout the observations. The presentation order of these scenarios was counter-balanced between groups. Similar to the design used in [10], 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. An overview of all tasks can be found in Table 1.

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 them 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. Both observers used notes and video samples to form initial coding categories. These were used by one observer for the first video coding pass and were refined through subsequent study of the videos and the second coding pass.

FINDINGS

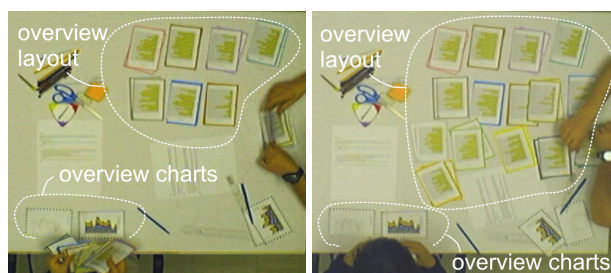
In this section, we outline our understanding of the collaborative and individual visual analysis process we uncovered during our analysis. We follow this by illustrating how the processes themselves were *not* temporally organized in a consistent way across groups. In the next section, we relate these 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 (summarized in Table 2). We describe each process using real 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 are an aggregation of several instances of particular processes during both scenarios.

Browse:

The browsing process comprises activities involving *scanning through data to get a feel for the available information*. Browsing activities do not involve a specific search related to a task; instead, the main goal is 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 11 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 ≈ 30 s for groups, and ≈ 60 s for individuals), perhaps indicating that, 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.



(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.

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. On average, pairs and triples spent 2.5 min reading and re-reading the task description; however, individuals referred to the task sheet more frequently (10 times vs. 9 times in pairs and 7 times for triples in total). 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:

Five teams explicitly discussed their *overall task division strategy*. We observed three main collaboration strategies:

- *Complete task division.* Participants divided tasks between themselves to avoid duplicating work. Each participant worked alone with his or her information artefacts on a separate subset of the problems. Results would be combined at the end without much further group validation.
- *Independent, parallel work.* Participants worked concurrently on the same tasks but independently of each other. When one participant had found an answer, solution and approach were compared and discussed. Other participants might then validate the solution by retracing the approach with their own artefacts, or by carefully examining a 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 (in terms of conversation and providing assistance) using primarily their own 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 8 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 brief: < 1 min on average per scenario. Most of the changes in task 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.

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

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 22 times with the help of individual information artefacts for all groups and tasks; 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).



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

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 resolved twice using other charts as aids. Others did not attempt a clarification but chose alternative representations leaving out the one that was unclear. In teams, the need for clarification involved discussion with other participants to decipher and understand the charts and sharing of information artefacts. Overall clarification required less than 1 min for Scenario B and no clarification was required for Scenario C. The clarification times for Scenario B were higher for each group 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 (1 min, 2 min) for clarification in Scenario B.

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.* Starting from a pile-based categorization of information artefacts, piles are scanned and relevant cards 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 were *relocated* and placed side-by-side for comparison. Figure 5 shows an example on the right where a variable was to be measured, so the card was *relocated* closer in the individual person’s workspace. The spatial organization of cards relative to piles of data could carry 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 re-placed. Similarly, the organization scheme might reflect the perceived importance of a set of cards: 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 ≈ 4 min selecting data, the second most common process in our study.

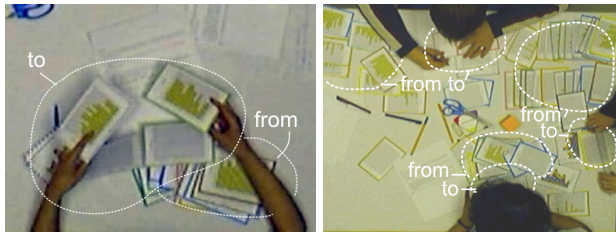


Figure 5. Chart organization during selection depending on their intended usage. Left: a participant selected four cards for comparison placing them side by side in her hand. **Right:** three participants selected individual charts and placed them in the center of their workspace to measure a specific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

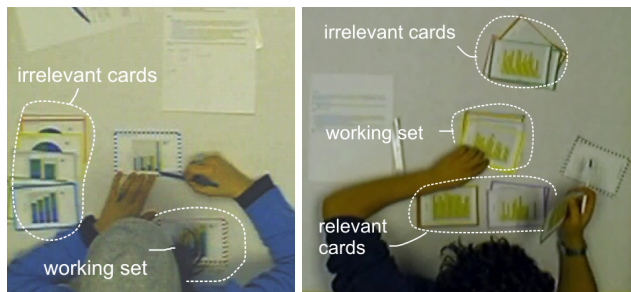


Figure 6. Changing categorization during selection. Left: a participant placed irrelevant cards to her left and picks single cards to operate on from the working set. **Right:** a participant picked out relevant cards, placed them close to himself, and put irrelevant cards in a pile further away.

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 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). Comparing values on a specific chart or values across charts was also extremely frequent. 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 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 min) on operations per scenario. 64% of operations 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: on 47 occasions participants validated others' solutions by looking carefully at the solution using shared representations, while other times they searched for 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 it. 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 (≈ 3 min), pairs spent ≈ 2 min 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 oc-

curred 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 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 $\approx 70\%$ 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 both charts 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 [10] 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 [6], or to ensure a correct or agreed-upon approach.

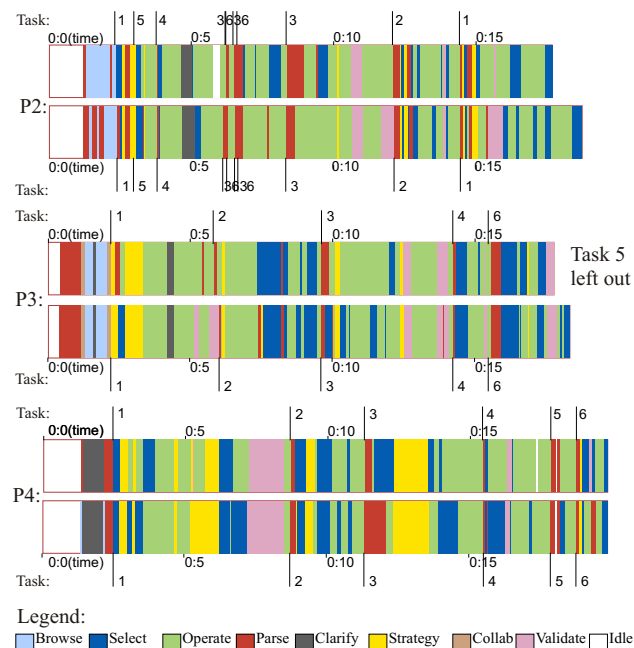


Figure 8. Temporal sequence of processes for three pairs during one complete scenario.

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.

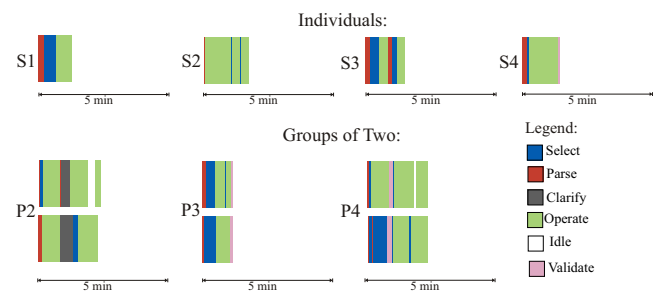


Figure 9. Temporal sequence of processes for one open discovery task. The top row shows timelines for individual participants (S1–S4). The bottom row holds timelines for participants in groups of two (P2–P4).

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. [13] which involved observations of collaborative work, and an extension can be found in [20].

The Sense-Making Cycle has several components related to our model. It outlines a process called “foraging for data” that includes our *browse* process. Spence [16] specifically explores 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. Spence distinguishes three different browsing activities [17]: *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 who 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 [15].

“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. “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 may 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. It makes a strong temporal (cyclical) suggestion but does allow for loops within this cycle over defined forward and backward connections between 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. An adaptation of the Sense-Making Cycle by Pirolli and Card is presented in [20] for some type of analysis work. This extension includes two main components: A Sense-Making Loop in which a mental model of the data is iteratively developed and a Data Foraging Loop in which information is searched, read, filtered, and extracted. This model tries to cover most aspects of intelligent analysis work and our processes mostly relate to those parts within the Sense-Making Loop as discussed above.

Collaborative Analysis Models

In studying pairs using distributed CAVE environments, Park et al. [12] articulate a five-stage pattern of behaviour: **prob-**

lem interpretation, agreement on vis tool to use, search for a trend, discovery reporting, and negotiation of discoveries. Mark et al. [10] 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.** 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 [10] focuses on a context where the pair negotiates exploration through a *shared* tool (i.e., they could not work in a decoupled fashion [18]) whereas Park et al.’s model [12] 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” [10] and Park et al.’s “problem interpretation” [12] 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 either model. 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 [18], 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 [10], like Park et al.’s “agreement on visualization tools to use” [12]. 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 [10]. 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 [10, 12]. 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 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. Similar observations have been made by Heiser et al. [7] in a study of non-digital co-located and remote sketching activities. 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 assess 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. For example, one person should be able to select data from or browse a database while another already works on previously selected information.

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:

Information artefacts were re-arranged on the table by all 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 [14] 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 the workspace. During *browsing*, overview layouts were often created in which the cards were spread across 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 [11] 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. We have also given concrete design implications for digital information visualization systems derived from our findings.

ACKNOWLEDGEMENTS

This work was supported in part by Alberta Ingenuity, iCORE, NSERC, NECTAR, and Smart Technologies. We would also like to thank Stacey Scott and our co-workers in the ilab for important feedback on this work and Bob Spence for discussions about navigation in information spaces.

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.
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. J. M. Carroll, M. B. Rosson, G. Convertino, and C. H. Ganoe. Awareness and Teamwork in Computer-Supported Collaborations. *Interacting with Computers*, 18(1):21–46, 2006.
5. 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.
6. H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, UK, 1996.
7. J. Heiser, B. Tversky, and M. Silverman. Sketches For and From Collaboration. In *Visual and spatial reasoning in design III*. Key Centre for Design Research, Sydney, 2004.
8. P. Isenberg and S. Carpendale. Interactive Tree Comparison for Co-located Collaborative Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1232–1238, Nov./Dec. 2007.
9. 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.
10. 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.
11. MayaViz. CoMotion®. Website, 2007. <http://www.mayaviz.com/> (accessed March 2007).
12. 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.
13. 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.
14. 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.
15. 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 Computer Society.
16. R. Spence. A Framework for Navigation. *International Journal of Human-Computer Studies*, 51(5):919–945, November 1999.
17. R. Spence. *Information Visualization*. Pearson Education Limited, Harlow, England, 2nd edition, 2007.
18. 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.
19. J. C. Tang. Findings from Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies*, 34(2):143–160, February 1991.
20. 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.
21. F. B. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. 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, Nov./Dec. 2007.