A Framework for Visual Information Analysis

Petra Neumann, Anthony Tang, Sheelagh Carpendale

Abstract—To design information visualization tools that support users' needs, we need to understand how users engage with information visualizations in their analysis process. With the rapid growth in size and complexity of datasets, the practicality of an individual analyzing an entire dataset is becoming unrealistic. Instead, the expertise to make informed decisions about these information-rich datasets is often best accomplished by a team. However, there exist relatively few models that describe the visual analysis process, and only few studies that explore the differences between how individuals and teams use visualizations. We present an observational study where we explored the information analysis process of groups and individuals in the context of visual information. From the analysis of our study, we derive a *framework* that captures the activities of co-located teams and individuals engaged in information analysis. This framework has implications for the design, heuristic evaluation, and analysis of both collaborative and single-user digital information visualization tools.

Index Terms—Information Visualization, Analysis Process, Collaboration, Evaluation.

1 Introduction

Interactive information visualization tools have fundamentally changed how we analyze and think about information by allowing us to manipulate views and representations of this information [15]. As a consequence, these visualizations are often the center of many complex information analysis tasks. Since these tools affect how we can think about data, it is imperative to design these tools well. Human-computer interaction (HCI) researchers have long recognized that to design effective tools to support the flow of work, it is important to understand the *cognitive processes* that are at work. Since we are interested in designing information visualization tools, we therefore also take an interest in understanding the nature of the *visual information analysis process*.

In everyday practice, data is often interpreted and analyzed not only by individuals but by *teams* of individuals working in concert to make decisions. How are information visualizations used by these teams? How *could* they use these information visualizations in their collaborative process? While many researchers have explored the information analysis process [4, 7, 14], little has emerged on the nature of this process in a collaborative context [8, 11]. In particular, we are interested in the differences between how individuals and small co-located teams (e. g., two to three individuals) make use of visual information in solving problems involving visual information. How a single doctor would analyze biomedical information visualizations, for example, might differ from how a team of doctors might analyze the same data.

To address this problem, we designed an observational study to understand the flow and nature of this collaborative process and its relation to individual analysis practices. Specifically, we focused on analyzing how participants engage with the workspace and their collaborators to derive practical guidelines for information visualization tool design. Teams in our study were given paper-based (static) visualizations to solve tasks, allowing us to view their process independently of the constraints of a specific information visualization 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 observational study aimed to study 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 five concrete design implications for digital information visualization tools derived from our findings.

- *Petra Neumann is with the University of Calgary, E-mail: pneumann@cpsc.ucalgary.ca.*
- *Anthony Tang is with the University of British Columbia, E-mail: tonyt@ece.ubc.ca.*
- *Sheelagh Carpendale is with the University of Calgary, E-mail: sheelagh@cpsc.ucalgary.ca.*

2 Related Work

We focus on the particular problem of understanding how individuals and teams solve tasks using information visualizations. In this section, we discuss two bodies of work that relate to this problem. First, we set the scene by outlining efforts to articulate the "information visualization process," or the process through which one extracts insight from a dataset given a problem and a visualization tool. These efforts informed our interpretations of study participants in our work. Second, we discuss two representative research efforts in the field of computer supported cooperative work (CSCW) that were able to uncover work practices that comprise collaborative activity, and discuss briefly how we drew from this approach in our work.

2.1 Understanding Visual Information Analysis

Several researchers have outlined frameworks that describe how individuals make use of information visualizations to solve problems. These frameworks share the common characteristic of modeling a user's involvement in the visualization process as an iterative sequence of components; however, each is unique in level of abstraction and focus. In this section, we outline several of these perspectives to build a foundation upon which we interpret the findings in our own study.

Card et al. [4, pp. 10] provide a high-level model of human activity called the "knowledge crystallization cycle" where the goal is to gain insights from data relative to some task. The components of this model include activities ranging from foraging of data to work on to deciding or acting on the findings. Spence [14] 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.

Other researchers are more concerned specifically with the design of digital information visualization tools, and focus on how users manipulate view and visualization transformation parameters [4, 5, 7]. For instance, Jankun-Kelly et al. propose a model of visual exploration for analyzing a user's interaction with a digital visualization system [7]. Their goal was to capture interactions with the system's visualizations, and to tie the visualization results to the controlled visualization parameters. The underlying proposition of this model is that the manipulation of visualization parameters in digital systems is a *fundamental operation* in the visual exploration process. These models are

effective in capturing the temporal aspects of visual parameter manipulation; however, they do not capture the higher-level semantics of a user's interaction (i. e., why did the user change that parameter).

How individuals work with information has also been looked at from a task-centric perspective. For example, Shneiderman outlines a two-step process ("overview then detail"), and suggests seven different tasks that information visualization tools should support in order to facilitate the problem solving process: overview, zoom, filter, details-on-demand, relate, history, and extract [13]. These tasks have been generally accepted as guidelines for the design of information visualization tools [6]. 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].

Yet how do these models apply in the context of *collaborative* visual information analysis? In studying pairs using distributed CAVE environments, Park et al. articulate a five-stage pattern of behaviour ranging from problem interpretation to negotiation of discoveries [11]. Mark et al. also provide a five-stage collaborative information visualization model (Figure 1) [8]. 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 last two models share some similarities, but are clearly not identical. We argue that these differences suggest that we have only begun to understand the collaborative visual analysis process. A possible explanation for the disparity is that Mark et al.'s model [8] focuses on a context where the pair negotiates exploration through a *shared* tool (i. e., they could not work in a decoupled fashion [17]) whereas Park et al.'s model [11] allows for more loosely coupled work.

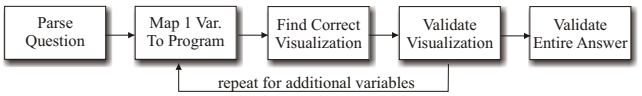


Fig. 1. Mark et al. [8] outline a five-stage model of collaborative information visualization tool use.

Although these models have not been developed in the particular co-located group and individual work context that we are interested in (Card et al.'s model focuses on a single user; Park et al. focus on immersive CAVE environments, and Mark et al. focus on pairs mediating their interaction with a single-user visualization environment). These perspectives on the information visualization process have informed the interpretation of our results. Specifically, we later revisit models by Card et al. [4], Mark et al. [8], and Park et al. [11] in the context to our findings.

2.2 Studying Work Practices through Observation

We are interested in studying how people solve visual analysis tasks as a group or as an individual in order to develop information visualization tools that can support this process. Researchers in the CSCW community frequently study how users accomplish tasks in *non-digital contexts* in order to understand what *digital tools* should support (e. g., [12, 18]). This approach generally relies on qualitative methods, including observation of users, inductive derivation of hypotheses via iterative data collection, analysis, and provisional verification [16]. In the CSCW literature, this style of research works well to uncover the basic mechanics of collaborative work. 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 [18]. 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 [12]. 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. Of particular interest is that in these studies, the researchers chose *not* to use digital tools, and instead to study the participants using traditional artefacts, such as pens, paper, cardboard, and so forth. The reasoning behind this choice is that participants' *physical interactions* with these familiar artefacts and tools would closely reflect how participants *understand and think* about the problem at hand.

Our work builds on this basic observational methodology in order to understand, based on participants' interactions with artefacts, the basic activities in the visual analysis process. This approach is not yet widespread in the information visualization community, which has largely focused on performance evaluations of the use of visualization tools; however, this approach is well-suited for our growing interest in understanding perception, exploration, and discovery in visualization systems.

3 An Observational Study of the Information Analysis Process

We conducted an observational study to understand the visual analysis process. We wanted to observe participants' natural working styles, unencumbered by any specific digital interface, so we developed a set of static charts placed on index cards to represent the visualization tool, and provided participants with traditional tools such as pens and paper. This setup allowed us to observe behaviours such as free arrangement of data, annotation practices, and different ways of working with individual information artefacts—behaviours that we would not otherwise see given most digital 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, like Mark et al. [8], our specific interest was in uncovering the *general processes* involved in collaborative and individual visual analysis.

3.1 Participants

We recruited 24 paid participants from the university population: 4 participated as individuals (1 female, 3 male), 8 participated in pairs (4 female, 4 male), and the remaining 12 participated in groups of three (9 female, 3 male). The mean age of the participants was 26 years. In total, we conducted twelve sessions with 4 individuals, 4 pairs, and 4 groups of three solving our information analysis tasks. We chose differently-sized groups (pairs vs. groups of three) to understand whether group size would have a noticeable impact on collaborative work practice, and also studied individuals as a baseline.

3.2 Design

Participants worked on two task scenarios with 4 and 6 tasks each. Each scenario contained a different set of data and representations in order to find out if they influenced how participants solved the tasks. The presentation order of these scenarios was alternated between groups. Similar to the design used in [8, 9], our scenarios each contained an equal number of *open discovery tasks*, where tasks could have several possible solutions, and *focused question tasks* that had only one correct answer. For example, one scenario contained study data on ratings of appropriateness of 15 behaviours in 15 different situations. In this scenario, an example of an open discovery task was, "choose three situations and describe behaviours most appropriate for that situation according to the graphs," and an example of a focused question was, "is it more appropriate to argue or belch in a park?" Each of the tasks could be solved using the data given to participants.

3.3 Apparatus

Participants were provided with a large table (90×150 cm) as their working environment. The table was covered with a large paper sheet, and several pens, pencils, rulers, erasers, scissors, and sticky notes were provided. Participants were given 15×10 cm cards with static charts. These charts showed different subsets of the data using at least two different representations (e. g., line chart and bar chart), and each group member was given a complete set of charts for each scenario. We refer later to these cards as *information artefacts* in the workspace.

3.4 Method

Participants were greeted and allowed to seat themselves around the table. Participants were given a short tutorial on the types of charts, tasks, and scenarios provided in the study. We told participants that they could use any of the tools (pens, rulers, etc.) to work with the graphs, 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 and to answer any questions. Once it was clear how to proceed, we gave participants each task scenario in turn, instructing them 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. We instructed single participants to use a "talk aloud" protocol. During each session two observers were always present. Both observers collected notes, and each session was video or audio taped.

3.5 Analysis

We analyzed field notes and video data using an open coding approach [16]. Initial coding categories were informed by our notes, and these coding categories were iteratively refined through further analysis of 610 minutes of video data (roughly 50 minutes for each session). This method provided a rich understanding of the similarities in the collaborative activity across groups, and the unique character of each group.

3.6 Findings

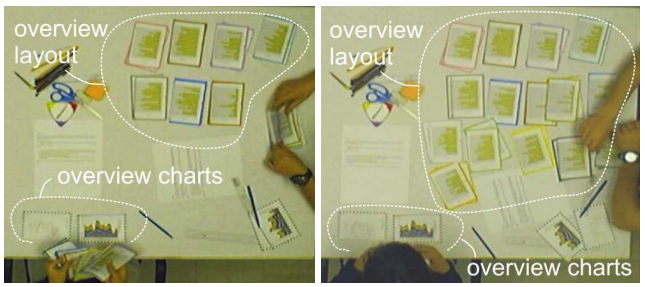
In this section, we outline our understanding of the collaborative and individual visual analysis process. We follow this by illustrating how the processes themselves were *not* temporally organized in a consistent way across groups and individuals, suggesting that information visualization tools should support this flexibility. In the next section, we relate this framework to prior work, and discussing how it can inform the design of information visualization tools.

3.6.1 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-specific strategy, clarify data, operate on data, select data, and validate findings (summarized in Table 1). We describe each process in this section using real examples drawn from our study, discussing participants' interactions with one another and the workspace. Participant group size did not meaningfully affect the nature of these processes; however, where individuals exhibited these processes differently, we describe these subtleties in the appropriate subsection.

Browse: The browsing process comprised activities involving *scanning through data to get a feel for the available information*. Browsing activities did not seem to involve a specific search related to a task, and appeared to be done mainly to examine or understand the data set. Specifically, these activities included quick glances or scans of the information artefacts to gain an understanding of both the types of charts available and the variables in the charts. Participants used several different techniques for browsing. Some participants took the complete pile of charts and flipped through them in their hands, while others created an elaborate layout of cards on the table. Figure 2 shows an example in which two participants use two very different browsing strategies. The participant on the bottom 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.

Parse: The parsing process captures *the reading and interpretation of a task description* with the goal of trying to understand how to solve the problem. While many real-world information analysis scenarios may not have a concrete problem description, an assessment of the given problem(s) and the required variables can certainly still occur and would be considered part of this process. When working in teams participants were each given a task description sheet. The reading of the task description occurred both quietly or aloud, and this reflected



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

Fig. 2. 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.

the collaboration style that teams adopted. When read aloud, for instance, collaborators could maintain a joint awareness of the state of the activity, and which problems various collaborators were working on: reading aloud was often followed by a discussion about what variables would be of interest, or how the variables should be examined. In both group and individual sessions, reading and interpreting the task sometimes resulted in a rephrasing of the question or note-taking of required variables. Parsing occurred frequently, not just at the beginning of the task, but multiple times—even for the same task: participants would often refer back to the problem sheet when they were uncertain of how to proceed, and would re-read the task description. Because the problem sheet would be referred to so frequently, it was treated as a special information artefact: it often had a prominent spot in a participant's workspace and was seldom moved. Figure 3 shows two examples of typical placements of the problem sheet in participants' workspaces. Even if the problem sheet was covered, as for two of the

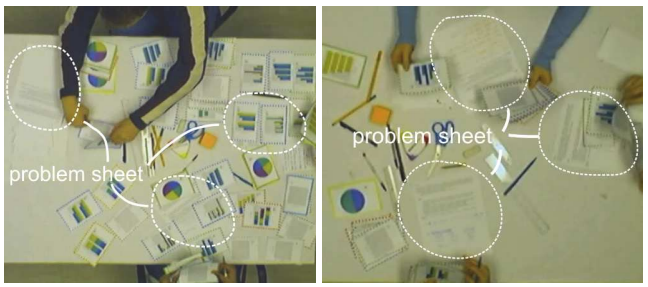


Fig. 3. Two typical examples illustrating how the problem sheet (outlined) received a prominent spot in participants' workspaces.

three participants on the left, the sheet would usually not be moved but accessed by moving artefacts that covered it. The problem sheet was also often used as the primary notepaper to record answers, reinterpretations of the questions, or to retain action lists (e. g., variables to look for in the 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 styles:

- *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 prespecified 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

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-Specific Strategy	establish how to solve a task using the given data and tools	determine required views or interactions for the task
Clarify	understand a visualization	avoid errors in reading the visualizations
Select	pick out visualizations relevant to a particular task	find relevant views and visualizations
Operate	higher-level cognitive work on specific view of the data	solve task or sub-task
Validate	confirm a partial or complete solution to a task	avoid errors in completing the task

Table 1. The eight processes in information analysis. "Discuss Collaboration Style" only applies to collaborative analysis scenarios.

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 in a fairly joint manner (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, seven out of the eight groups *changed their collaboration strategy* midway through a task scenario or between scenarios. These changes most frequently occurred between independent and joint work on the tasks. In one instance, a team switched from complete division of tasks to joint work because one of the collaborators had already finished the tasks assigned to him. Six of the eight groups started with a loose definition of doing the tasks "together" which resulted in their switching frequently between independent, simultaneous work and joint work on the tasks. Most of these changes were quite seamless, and did not require any formal re-negotiation.

Establish Task-Specific 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 is to determine the next views or interactions required to extract variables or patterns from the data to solve the problem. As a group 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. To facilitate discussion, individual participants often re-engaged in a *parsing* process, re-reading the task description, and in Figure 4, the participants keep the problem sheets close at hand for this purpose.

This explicit strategy discussion was common in groups that 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." 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 2).

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

Fig. 4. Discussing a strategy on how to solve a task using the chosen chart. Notice that information artefacts are used as aids.

of the scenario or task descriptions, through annotations on information artefacts, and in a few cases, the drawing of example diagrams. In group settings, the need for clarification additionally involved discussion with other participants to decipher and understand the charts and sharing of information artefacts.

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 2), 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., highly 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 they deemed relevant in a single pass and then *operated* on them afterwards (see Figure 6, right).

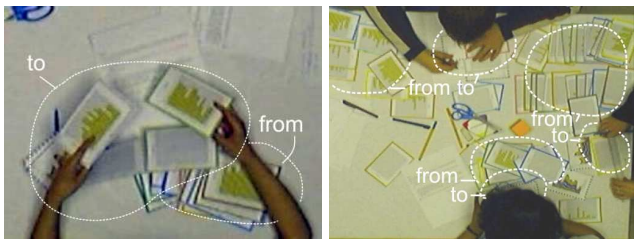


Fig. 5. Chart organization during selection depending on their intended usage. *Left*: a participant selected four cards for comparison. They are placed side by side in her hand. *Right*: three participants selected individual charts and placed them in the center of their workspace in order to measure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

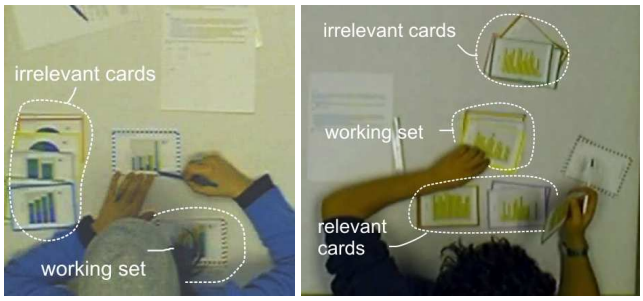


Fig. 6. Changing categorization during selection. *Left*: a participant placed irrelevant cards in an organization to her left and picks single cards to operate on from the working set. *Right*: a participant picked out relevant cards, placed them close to himself, and put irrelevant cards in a pile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

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. Comparing values on a specific chart or values across charts was also extremely frequent. In our study, participants usually 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 overlaying 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 group if other tasks depended on these results (e. g., during joint activity).

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 groups, 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

Fig. 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.

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 groups also rely on a collective validation from the social group.

3.6.2 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., *parse* frequently appears before *select*), *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 and clarity, we present only charts for individuals and pairs in this section; however, groups of three exhibited behaviour similar to pairs and individuals in that there was no consistent temporal relationship between the processes.

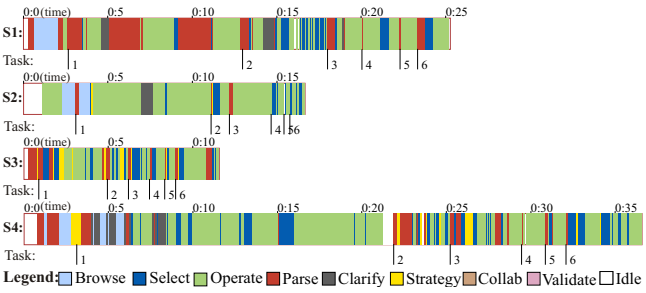


Fig. 8. Temporal sequence of processes for four individual participants during one complete set of scenario tasks.

Figure 8 shows the coded temporal sequence of analytic processes during an entire scenario for the four single participants. Notice how the sequence of processes was quite different for each participant, even though participants worked on the same tasks using the same tools, representations, and views of the data. Similarly, groups exhibited very unique temporal sequences of processes for solving the tasks. Figure 9 shows the working styles of three pairs (the remaining participant pair declined to be videotaped) during the same scenario as in Figure 8.

In both charts Tasks 1–3 were open discovery tasks and Tasks 4–6 were focused question tasks. We noticed that both individuals and groups solved focused question problems quicker than open discovery tasks. In general, groups spent a long time establishing a shared, common understanding of the problem during *parsing*, and spent a much

longer time than individuals *validating* their solutions. As a result groups had a better understanding of the tasks and solved them (both focused and open discovery tasks) more correctly. This result echoes findings in [9] that suggest that groups perform more accurately, albeit slower. Of course, groups also exhibit *establishing a task-specific strategy* more so than individuals, again in order to establish common ground, or to ensure a correct or agreed-upon approach.

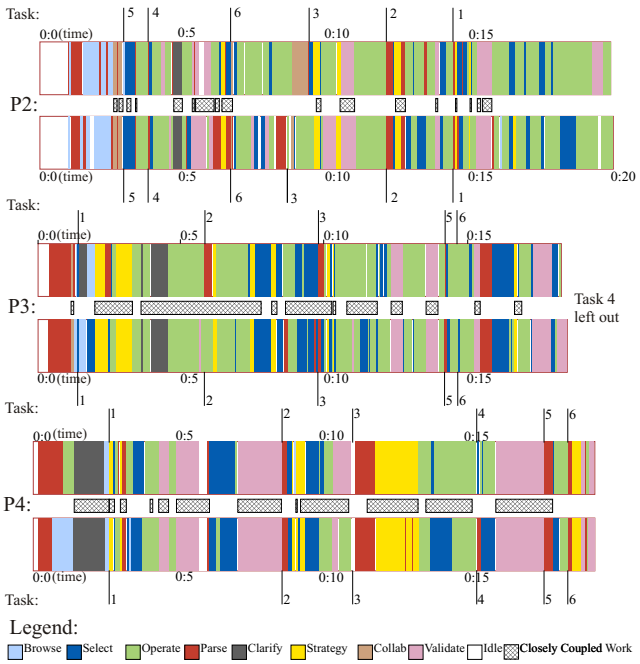


Fig. 9. Temporal sequence of processes for three pairs during one complete scenario.

Figure 10 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 widely. Phases of closely-coupled work include those in which participants joined in a discussion or shared artefacts in the workspace.

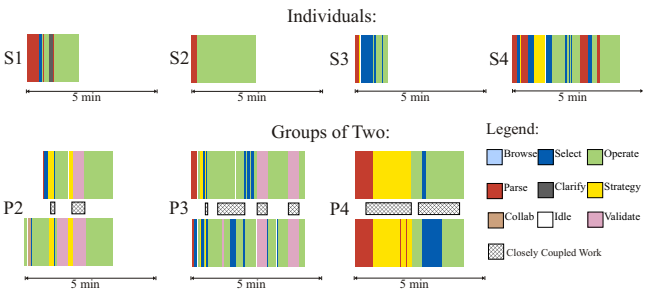


Fig. 10. Temporal sequence of processes for one open discovery task. The top row shows timelines for individual participants (S1–S4). The bottom row holds timelines for participants in groups of two (P1–P3).

4 Discussion: A Framework for Visual Information Analysis

To this point, we have introduced a set of processes that occur within the context of collaborative and individual visual information analysis. This eight-process framework, derived from our observational study, is unique from prior work in that it provides us with a way of understanding how *groups* and *individuals* use information artefacts in the workspace to solve visual information analysis tasks and how team members engage with each other in this process. In this section, we illustrate the validity of our framework by discussing how it relates to the information analysis/information visualization models from Section 2.1. This discussion will reveal that while individual processes relate closely to existing models, our temporal analysis suggests that with appropriate tools, both the collaborative and individual information analysis process can occur more fluidly without a pre-specified temporal order.

4.1 Relating the Framework to Other Models

Figure 11 provides an overview of the eight processes we derived from our user observation, and relates them to components of models by Card et al. [4], Mark et al. [8], and Park et al. [11]. In general, our analytic framework encompasses these models (and introduces two new processes: *clarification*, and *discuss collaboration style*). In this section we discuss each of our processes and relate them to the components of other models.

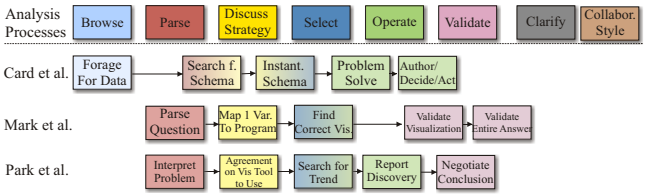


Fig. 11. The eight identified processes of information analysis in relation to the models by Card et al. [4], Mark et al. [8], and Park et al. [11].

Browse: Card et al.'s model outlines a process called *foraging for data* that relates to the browse process [4]. Spence further distinguishes three different browsing activities [15]: *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. Accordingly, 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 [13].

Parse: Our parsing process relates closely to Mark et al.'s *parse question* [8] and Park et al.'s *problem interpretation* [11] stages. Card et al.'s *search for schema* also seems to involve activities that we characterize as being a part of parsing, specifically the identification of attributes on which to operate later [4]. The activities of reading, parsing into distinct variables, and interpretation described in these models are augmented in our parse component by additional activities of discussion, and note taking found during our study.

Discussing Collaboration Style: Previous models do not discuss this process explicitly, but 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. Park et al.'s study results reflect this tendency for individual work using a localized view of the data set [11]. Similar differences in work styles for spatially fixed information visualization tasks (e. g., maps that cover the whole workspace) have been described in [17], but they have not been put in a greater context of other processes of visual analysis.

Establish Task-Specific Strategy: This type of planning is typically described from a tool-specific perspective. 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 [4]. According to Mark et al.'s model, *map 1 variable to program* is most closely related in that it would also involve a collaborative agreement on the most appropriate visualizations, parameters, or views to solve the problem [8], like Park et al.'s *agreement on visualization tools to use* [11]. Our description of this process discusses the activities involved in establishing a strategy rather than describing it in the context of a specific tool.

Clarification This process is unique to our framework, and 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.

Selection Our articulation of the *selection* process is related to parts of the activities covered by Mark et al.'s *find correct visualization* stage [8], Park et al.'s *search for trend* [11], and Card et al.'s *instantiate schema* [4]. Our description of *selection*, however, more broadly captures the notion of picking out important information beyond operations in a specific visualization system.

Operation All three previous 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 [3], *independent search for a trend including some adjustments to viewing parameters*, or *report discovery*. Operation is not an individual stage in Mark et al.'s model but is integrated in the *find correct visualization* stage [8].

Validation Validation is not directly represented in Card et al.'s model [4]. Perhaps, as we have also observed, because validation seemed to be often omitted or quite brief for individual participants. In groups, this stage was much more visible and it is also included in the pipelines by Mark et al. and Park et al. as the last stage of information analysis [8, 11]. 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.

4.2 Temporality and Process-Free Tools

Many of the existing models suggest a *typical* temporal order of components (Figure 11); however, our analysis of the temporal occurence of the framework processes in our study suggests that this typical temporal ordering was not particularly evident (Section 3.6.2). 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 maps clearly 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.

5 Implications for the Design of Information Visualization Systems

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 findings from our study.

Support Flexible Temporal Sequence of Work Processes: Individuals have unique information analysis practices based on their prior experiences, success, 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 in the way they force their users to work. 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 *digital systems should provide for a flexible order of operations to be performed*. Co-located collaborative systems, in which more than one user may work and interact at the same time, should *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 must be able to perform individual operations on these views unaffected by his or her team members' actions. However, the tool must 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. We refer to guidelines for the design of multiple views in information visualization for suggestions on addressing some of these issues [2].

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 [12] that call for co-located collaborative systems to *provide appropriate functionality in these personal workspaces* (territo-

ries). 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 [10] 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.

Support Flexible use of Representations: We found that participants quite frequently used different representations when they could not solve the task using one, if they found one too difficult to read, or if they wanted to validate an answer. Some participants showed clear preferences for different representation types. These preferences could vary within a group, session, and depending on the tasks. To support analysis of data, information visualization systems should *provide each person or group member with individual access to different types of representations*. Zhang and Norman found that providing different representations of the same information to individuals provide different task efficiencies, task complexities, and change decision-making strategies [19]. Allowing users to freely switch between different representations should be supported so that they do not lose their current focus of attention or their history of previous interactions with a representation. For example, annotations should be made transferable between representations, a feature that is not frequently supported.

Support Flexible Interaction with Information Artefacts: We identified two common types of activities during the operate process in our task context: extracting a single data value and comparison. Extracting of data values required mostly sequential access to representations while comparison required simultaneous access to at least two different representations or views. While these activities were specific to our tasks, it seems clear that *providing simultaneous access to different datasets, representations, or views* can be important to support the operate process.

Most of our processes contained some form of notetaking or annotation activity. Note-taking mostly occurred during parsing or to report a result at the end of an operation. During the course of both scenarios each participant on average annotated at least three information artefacts. *Supporting free annotations and note-taking* capabilities can enhance a user's thought processes during analysis.

6 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*. 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] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for Using Multiple Views in Information Visualization. In *Proc. of the Conference on Advanced Visual Interfaces (AVI)*, pages 110–119, New York, USA, 2000. ACM Press.

[3] J. Bertin. *Semiology of Graphics: Diagrams Networks Maps*. The University of Wisconsin Press, Madison, WI, USA, 1983. Translation of: Sémiologie graphique, 1918.

[4] 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.

[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] B. Craft and P. Cairns. Beyond Guidelines: What can we learn from the Information Seeking Mantra? In *Proc. of the Conference on Information Visualization (IV)*, pages 110–118, Los Alamitos, USA, 2005. IEEE Comp. Society.

[7] 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.

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

[9] 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 Conference on Information Visualization (IV)*, pages 249–255, Los Alamitos, USA, 2002. IEEE Comp. Society.

[10] MayaViz. CoMotion Discovery. Website, Accessed April 2006. http://www.mayaviz.com/.

[11] 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.

[12] 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.

[13] 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.

[14] R. Spence. A Framework for Navigation. *International Journal of Human-Computer Studies*, 51(5):919–945, November 1999.

[15] R. Spence. *Information Visualization*. Pearson Education Limited, Harlow, England, 2nd edition, 2007.

[16] A. Strauss and J. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Thousand Oaks, London, New Delhi, 2nd edition, 1998.

[17] 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.

[18] J. C. Tang. Findings from Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies*, 34(2):143–160, February 1991.

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