

Awareness and distraction in loosely-coupled collaborative brushing and linking

by

Amir Hossein Hajizadeh

B.Sc., Sharif University of Technology, 2010

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

in the Department of Computer Science

© Amir Hossein Hajizadeh, 2013
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by
photocopying or other means, without the permission of the author.

Awareness and distraction in loosely-coupled collaborative brushing and linking

by

Amir Hossein Hajizadeh

B.Sc., Sharif University of Technology, 2010

Supervisory Committee

Dr. Melanie Tory, Supervisor
(Department of Computer Science)

Dr. Daniela Damian, Departmental Member
(Department of Computer Science)

Supervisory Committee

Dr. Melanie Tory, Supervisor
(Department of Computer Science)

Dr. Daniela Damian, Departmental Member
(Department of Computer Science)

ABSTRACT

Maintaining an awareness of collaborators' actions is critical during collaborative work, including during collaborative visualization activities. Particularly when collaborators are located at a distance, it is important to know what everyone is working on in order to avoid duplication of effort, share relevant results in a timely manner and build upon each other's results. Can a person's brushing actions provide an indication of their queries and interests in a data set? Can these actions be revealed to a collaborator without substantially disrupting their own independent work? I designed a study to answer these questions in the context of distributed collaborative visualization of tabular data. Participants in my study worked independently to answer questions about a tabular data set, while simultaneously viewing brushing actions of a fictitious collaborator, shown directly within a shared workspace. I compared three methods of presenting the collaborator's actions: brushing & linking (i.e. highlighting exactly what the collaborator would see), selection (i.e. showing only a selected item), and persistent selection (i.e. showing only selected items but having them persist for some time). My results demonstrated that persistent selection enabled some awareness of the collaborator's activities while causing minimal interference with independent work. Other techniques were less effective at providing awareness, and brushing & linking caused substantial interference. These findings suggest promise for the idea of exploiting natural brushing actions to provide awareness in collaborative work.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	iv
List of Figures	vi
Acknowledgements	vii
Dedication	viii
1 Introduction	1
2 Related Work	7
2.1 Awareness in Collaborative Work	7
2.2 Attentionally Ambient Visualizations	9
3 Operam: Synchronous Distributed Collaboration	10
3.1 Concepts	11
3.2 Personal Use	12
3.3 Collaborative Use	15
4 Study Design	17
4.1 Participants	18
4.2 Apparatus	18
4.3 Tasks	19
4.3.1 Primary task (data analysis)	19
4.3.2 Secondary task (awareness)	19
4.4 Visualization Conditions	20

4.4.1	Brushing & linking	20
4.4.2	Selection	21
4.4.3	Persistent selection	21
4.5	Procedure	22
4.6	Hypotheses	23
5	Results	25
5.1	Awareness of Collaborator's Actions	25
5.2	Primary Task Completion Time	26
5.3	Primary Task Accuracy	27
5.4	Perceived Distraction	28
5.5	Qualitative Comments	29
6	Discussion	31
6.1	Discussion of Hypotheses	31
6.2	Implications for Design	33
6.3	Design Details for Further Study	34
6.4	Limitations and Future Work	35
6.5	Threats to validity	36
6.5.1	Internal validity	37
6.5.2	External validity	37
6.5.3	Construct validity	38
7	Conclusions	39
A	Study Details	40
A.1	Questions	40
A.1.1	Answered by participant	40
A.1.2	Answered by fictitious user	40
A.2	Pre-questionnaire	41
A.3	Post-questionnaire	41
	Bibliography	42

List of Figures

Figure 1.1 Example of brushing and linking	3
Figure 1.2 Operam prototype screen shot	4
Figure 1.3 Collaborative conditions	5
Figure 3.1 Operam workspace view	12
Figure 3.2 Operam login page	13
Figure 3.3 Operam chart management	14
Figure 3.4 Brushing and linking in Operam	15
Figure 5.1 Graph of correctly identified questions	26
Figure 5.2 Graph of completion time	27
Figure 5.3 Graph of accuracy	28
Figure 5.4 Graph of distraction	29

ACKNOWLEDGEMENTS

I would like to thank:

My friends, especially Nazpar Yazdanfar and my colleagues in VisID group,
for supporting me in the low moments.

My supervisor, for mentoring, support, encouragement, and patience.

and SAP for funding this research and providing me with the chance of an internship, from which I learned a lot.

DEDICATION

I dedicate this thesis to my parents who supported me even from miles away.

Chapter 1

Introduction

In visual analytics, we aim to support the entire process surrounding data analysis and sensemaking. This process typically includes not just interacting with data visualizations, but also recording notes and findings, synthesizing hypotheses and evidence, documenting findings for dissemination, and frequently, collaborating with other analysts. Evidence has shown that some visualization tasks can be completed faster and with greater accuracy in collaborative settings [13, 16, 19, 20]. For example, Mark et al. [20] found that groups were more accurate than individuals at searching large datasets and free data discovery tasks. Analysts often work in teams, where team members with different experience and expertise work together to solve challenging problems. It is therefore critical to support the mechanics of collaboration along with other aspects of the data analysis process. I argue that the approaches needed to support collaborative visual analytics are somehow different from other types of collaboration, because people need to understand their collaborators' actions in relation to the data being investigated.

One critical aspect of an effective collaboration is awareness (i.e. keeping track of what one's team mates are working on, what insights they have found, and what avenues of investigation they have pursued). Awareness helps team members to share information with each other at appropriate times, and avoid redundant work, thereby improving overall team performance [13]. I propose to exploit people's natural brushing actions (a common interaction technique when working with a visualization containing linked views [18]) as one mechanism of providing awareness to a collaborator. I present results of an experimental study in which one user's brushing actions were presented to a remote collaborator. I investigate the extent to which these awareness cues interfere with individual work, and whether they can provide awareness of

another person’s activities.

Awareness is important in all forms of collaborative work, but is especially challenging to support during synchronous distributed collaboration (i.e. same time, different place). In synchronous collocated (i.e. same place) collaboration, there are some inherent awareness mechanisms. A person can look over a collaborator’s shoulder to see what the other person is doing, can observe gestures and physical actions, and can talk to inform their team-mates of how they are progressing in their work.

In distributed collaboration though, many of these awareness cues disappear, and it becomes more important to provide explicit awareness support through software. I presume that collaborators may be connected through audio and video links, and may have a shared screen. These forms of support may be sufficient for closely coupled work, when collaborators are working closely together (e.g. discussing their findings or planning the next step). However, they may be insufficient during periods of loosely coupled collaboration (e.g. where participants have divided up the work and spend an extended period working independently). It is well known that same-time collaborative work naturally transitions between close and loose coupling [27], and that tools need to support many such collaboration styles and the transitions between them. One way to support loosely coupled work styles, plus timely transitions to more closely coupled work styles, is to help people maintain an awareness of what their collaborators are working on [16].

I focus on enhancing awareness during synchronous but loosely coupled work by distributed teams. I view this as the most challenging situation in terms of awareness, and therefore the most important to support through software. Many mechanisms designed to increase awareness between users use extra screen real-estate or will result in a cluttered screen that can interfere with user’s work and distract them from their primary task. Any visual representation of what the other person is doing will add more information to a potentially busy screen. Such representations could be distracting and may also require users to learn a new visual encoding. I aim to develop an awareness mechanism that (1) is effective at indicating to an information “consumer” what a “producer” is investigating, (2) exploits people’s natural interactions with a visualization tool (i.e. requires no extra effort on the part of the producer), (3) re-uses existing visual encodings within the tool so that no new encodings need to be learned, and (4) causes minimal interference with the consumer’s own independent work.

Brushing and linking is a mechanism commonly used in visualization tools to help

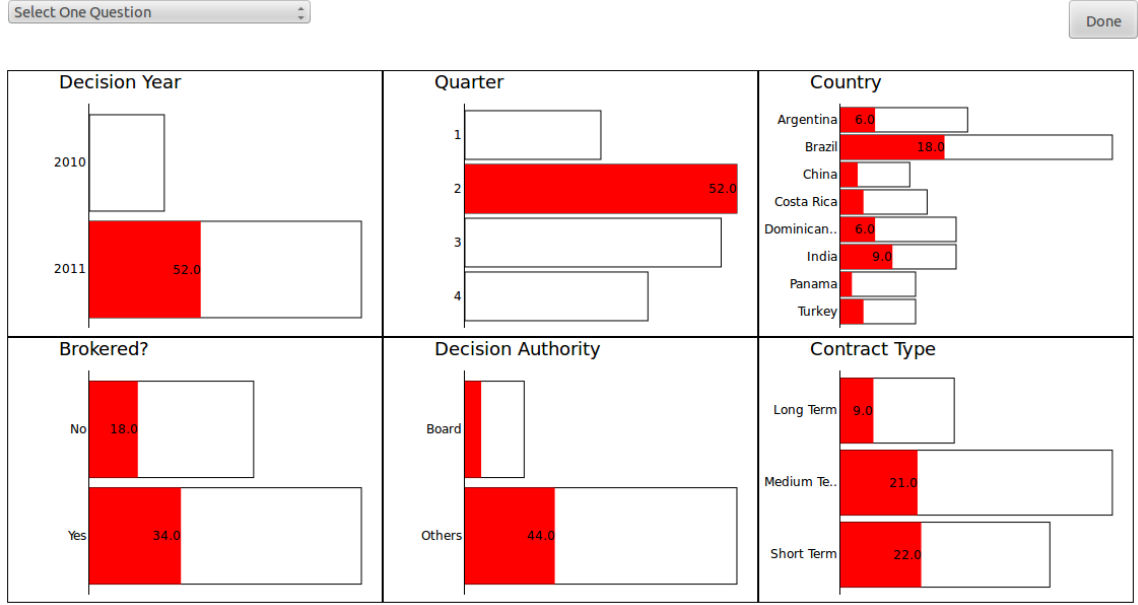


Figure 1.1: Example of brushing and linking in the charts used for my study. Quarter 2 has been selected by the analyst in the top middle chart. Data corresponding to Quarter 2 are highlighted in all other charts. Partial bar highlighting indicates the proportion of the data that match Quarter 2.

people explore relationships between data in different but related views [18]. Brushing lets the user select some data on one chart and see how it is reflected on other linked charts, as illustrated in Figure 1.1. For example, a user may select a cluster of people with high life expectancy, and see how this subset of data distributes over gender or country of origin. Brushing is an extremely common interaction technique in modern visualization tools for tabular data, and an effective strategy for understanding relationships between data subsets and dimensions. The data items that a person chooses to brush should be strongly related to the person’s questions and interests. Therefore, I suspected that people should be able to gain some knowledge of each other’s interests and investigative questions based on seeing brushing actions. Showing brushing actions to a distributed collaborator may therefore provide a useful awareness cue.

I aim to answer the following research questions: (1) Can an individual gain awareness of a remote collaborator’s activities by viewing a representation of their brushing actions? and (2) What visual representation of brushing actions will maximize awareness while minimizing interference with individual work? This study only focuses on the loosely coupled phase of the collaboration and is only one step in inves-



Figure 1.2: Screen shot of a distributed collaboration prototype I developed. Each user has their own work area where they can create charts. Shared parts of collaborators' workspaces may be viewed by selecting the appropriate tab.

tigating if users can gain awareness viewing brushing actions of a collaborator. More experiments are required to extend this research to other more realistic situations and to explore the effect on the transition between loosely coupled and closely coupled states.

My research questions arose from a design challenge I encountered while developing a visualization tool for synchronous distributed (same time, different place) collaboration (Figure 1.2). This tool was called Operam. The goal with Operam was to explore design options to support collaborative business intelligence. In particular, Operam supported collaborative analysis of business data (e.g. sales data tables) by enabling analysts to create and examine multiple linked charts. Operam is explained more in detail in chapter 3.

While it was fairly straightforward to design shared and individual workspaces to support both closely coupled and loosely coupled activities, it was less clear how brushing should be handled. Should collaborators be able to see each other's brushing actions? If so, when? Should all details be shown or just limited detail? During closely coupled collaboration, it seemed clear that having an identical shared view (including brushing) could help collaborators to discuss and analyze data together. This is

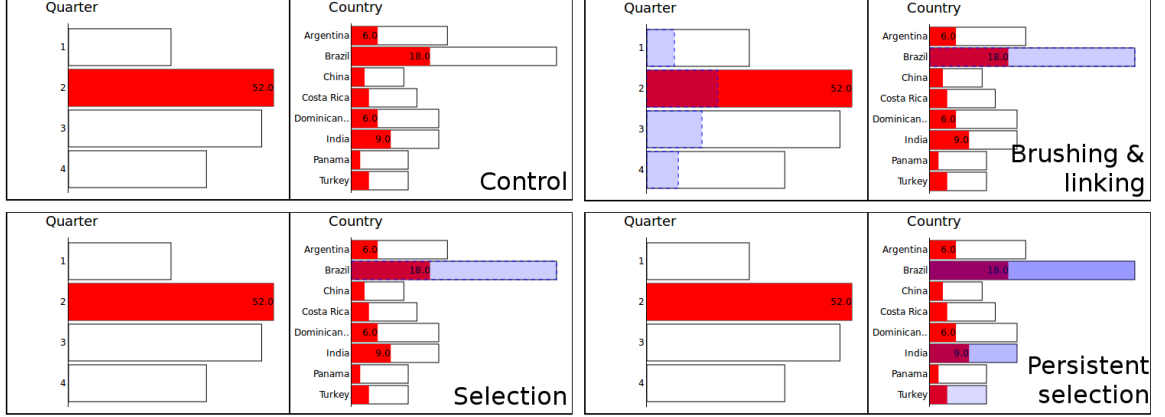


Figure 1.3: Four collaborative brushing and linking conditions used in my study. In the study, participants saw six linked charts, but here only two are shown for simplicity. Red = own brushing, Blue = collaborator. (Top left) Control condition with no awareness mechanism, (Top right) Brushing & linking, (Bottom left) Selection, (Bottom right) Persistent selection.

supported by previous work indicating that shared visualizations enable people to ground their discussions [13], but the best design for loosely coupled collaboration was not obvious. Seeing another person’s brushing actions could provide some awareness of what the other person is working on, but might also be distracting or otherwise interfere with individual work. What was the best approach to take?

I note that the general idea of collaborative brushing and linking has been previously introduced by Isenberg et al. [15]. This earlier work was inspirational to me and encouraged me to consider the possible value of collaborative brushing and linking. However, substantial differences in the task and data type made it non-obvious how to implement collaborative brushing in my situation. In particular, Isenberg et al. designed a tool for collaborative analysis of document collections rather than tabular data. In their work, “collaborative brushing and linking” was a feature explicitly added to the design for the purpose of supporting awareness. A visual indicator was shown to a collaborator when there was a common search term or a common document had been viewed or returned as part of a search. Individual user activities consisted of searching for and reading documents. In contrast, in my work, brushing and linking is the primary form of interaction for investigating data. I propose to further exploit those interactions for the secondary purpose of providing awareness.

I conducted an experimental study to compare three different methods of showing another person’s brushing actions, as shown in Figure 1.3. Brushing & linking

(top right) shows exactly what the collaborator sees while they brush the data (the selection plus corresponding items in all other views). Selection (Bottom left) shows only the collaborator's selected item. Persistent selection (Bottom right) is the same as selection, except that selected items remain highlighted for some time (40 seconds in our study) and gradually fade; thus multiple successive selections may be visible simultaneously.

My results showed that persistent selection was the most effective awareness mechanism, and that it can be implemented in a way that causes minimal interference with individual work. Contributions of this work include:

- A better understanding of brushing & linking as an awareness mechanism.
- Assessment of the value of selectivity and persistence in awareness mechanisms. In particular, studying the effect on the level of awareness and amount of interference with individual work.

In the following chapters, a report of literature is presented (chapter 2), then design of the prototype inspired this research, Operam, is explained in chapter 3. Chapter 4 belongs to the main usability experiment performed. Results are explained in chapter 5 and are discussed in chapter 6. Finally in chapter 7 the conclusion of the thesis is presented. More details of the study are also available in appendix A.

Chapter 2

Related Work

This research has two aspects, awareness and distraction. In this chapter, literature background for these two areas is presented. The first aspect attempts to characterize awareness in collaboration and provides tools for supporting it.

Since the studied visualization methods can also be categorized as ambient visualizations, a review of ambient visualization methods and experiments is explained as the second aspect of this literature review.

2.1 Awareness in Collaborative Work

Importance of establishing common ground between members of a collaborating team has been acknowledged in several areas, such as visual analytics [6, 7] and linguistics [8]. It is well known that common ground is a requirement for successful communication [6] and that even an imperfect shared knowledge will reduce cost of the communication between collaborators [13]. Weaver [28] mentioned that maintaining a visual common ground is the most difficult problem in collaborative visualization.

Heer and Agrawala [13] emphasized that the awareness of others' activities is an important source for common ground in collaborative work. Such awareness includes knowledge of what has been done and what are the possible next steps. Isenberg [16] found that collaborative brushing and linking as an awareness mechanism can avoid redundant work and empower analysts to validate and build upon each other's results.

In the context of collaborative visualization, awareness has been provided using a wide range of possible designs. At least four different categories of information can be

shared among collaborators; what has been done, who has done it, when was it done, and why [26]. Based on the context and scenario a subset of these four categories will be presented to the user. There are also at least four different approaches to presenting awareness information. In some cases the indications are directly encoded in the shared workspace, e.g. using color to indicate changes [11, 15, 26]. Tam et al [26] also used text labels to represent type of action, while different colors were used to show “who” had changed it. Gross [11] used color darkness as the indication for the change, which faded away over time. This approach is similar to what I used in my persistent selection condition.

As a different approach, user activities can be aggregated and abstracted [13]. Brush et al. [4] have studied use of notification as an example of this approach. In some other designs a separate space is allocated to the awareness mechanism, such as timeline [9], or user activities list [14]. Finally, some awareness approaches rely on active user participation, for example through chat or comment threads [14] or annotations [17].

Awareness mechanisms have clear benefits for collaboration, but they do not come without costs. Grudin [12] states that every additional effort put into a collaborative tool should have direct values for its users. Also some studies (e.g. [10]) have pointed to the cost introduced to collaborative effort by the communication media, including the means provided to keep participants awareness of each others’ progress. McCrickard et al. [22] proposed a comprehensive model to understand benefits and costs in designing notifications and ambient systems. Their model tries to incorporate possible design goals (interruption, reaction and comprehension) in a successful system design. In my approach, there is no extra effort expended to producing the information since the brushing actions are part of the normal analysis process. However, there is potential for the visualization of those actions to be disruptive to collaborators in addition to providing awareness benefits.

Another potential cost of awareness mechanisms is screen real estate devoted to the awareness tool. Some evidence indicates that people are willing to sacrifice screen real estate for beneficial information, providing the information is useful and there is sufficient real estate to begin with. For example, a field study by Cadiz et al. [5] on a peripheral awareness interface indicated that people were willing to give up a portion of their screen to stay aware of important information related to daily tasks.

2.2 Attentionally Ambient Visualizations

My visualizations of a collaborator’s brushing actions are attentionally ambient; in other words, they are intentionally designed to be able to move in and out of a viewer’s attentional focus (e.g., by forming a secondary visual layer so that they can be ignored if desired). Pousman and Stasko define ambient information systems by several criteria: important but not critical information; can move from the periphery to the focus of attention and back again; subtle changes to reflect updates in information (should not be distracting); and are aesthetically pleasing and “environmentally appropriate.” [23, 24]. Attentionally ambient visualizations extend the idea of ambient visualizations to an attention-based space rather than a location-based one. Closely related to ambient visualizations are peripheral visualizations (i.e. visualizations designed to be in one’s peripheral vision).

As an example of early work in peripheral visualization, Bartram et al. [2] studied motion as visual encoding technique. They focused on how different motion parameters influence detectability and attentional distraction from a separate primary task. Results indicated that a slow linear motion can be a good compromise, with low distraction and high detection rate. These results suggested to us that showing a full representation of one’s brushing and linking actions to a collaborator could be quite distracting since it may involve many rapid motion events; this might cause interference with primary task performance.

Later, Somervell et al. [25], conducted a similar study. They showed that density in peripheral visualizations and an attentionally demanding secondary task both have negative impacts on performance at a primary task. They also concluded that longer-lasting peripheral visualization are easier to interpret, which is in line with [2] and the hypotheses for my study. Persistence also improves comprehension and memorability of the awareness visualization [21].

Chapter 3

Operam: Synchronous Distributed Collaboration

The research problem arose while I was designing a synchronous distributed collaboration tool, called Operam. In this chapter I talk about the design process of Operam and the questions I tried to answer along the way while designing this tool.

Operam was designed to help a group of colocated or distributed analysts to work with data by creating custom charts semantically linked with each other, share their findings by sharing the visualization and add permanent or temporary annotations to the charts. Operam can be used in a situation where two analysts talking over phone can share their findings, discuss some of the charts and even “point” to the charts during their discussion. It can also be used during individual effort to analyze the data, but in this chapter I will focus on its collaborative use.

Pointing to an artifact is a common interaction during collaboration. Using this interaction, one can attract their collaborators’ attention to a specific item and add comments or propose questions about it. “Pointing” can be a general action of directing one’s attention to, or putting emphasis on some object or fact, by any form of action from extending a finger to adding a highlight by software means. This action can be directed to a physical object, a virtual item on computer screen or an abstract concept.

Operam supports two kinds of pointing. Users can point to a chart or to a data item. The latter will highlight any representation of that data item. A more generalized definition of this mechanism is called “brushing”, which is the main topic of this thesis. Using Operam collaborators can discuss data items while looking at

different charts representing different aspects of the data. For example, first user can point to a certain kind of T-shirt from products of a fashion company, while looking at the amount of benefit gained from selling the item. At the same time the other user might be looking at the chart visualizing cost of producing each item, but the T-shirt being discussed will be highlighted on the chart, helping them maintaining the context of their discussion even while working independently.

However, since Operam is a collaborative tool users can manipulate or edit charts synchronously. What should happen if each user selects a different item? Should the latest selection overwrite earlier selections? Should we color-code the selections with respect to their actor? I tried to answer these questions in my later prototype which was used in a usability experiment, but before getting to that I will go through the concepts and design decisions around Operam.

3.1 Concepts

One fundamental concept in Operam is the workspace (Figure 3.1(a)). Each user logs into their personal workspace. Workspaces act like a container for up to six charts, with customized access level for the users. Number of charts was an arbitrary number, chosen to leave enough space for each chart, without the need for the user to scroll through pages. There are two types of workspaces, “personal workspace” in which only the owner can manipulate the charts, and “shared workspace” where all users can work synchronously. I acknowledge that during the collaboration users will want to work closely or loosely connected, so a collaborative tool should support users during both these stages. During loose collaboration users will work on their personal workspace independently and may not want to share their charts unless they find a valuable finding. In contrast, during tight or close collaboration having a shared workspace will help them discuss data and build common ground.

Although only the owner of a private workspace can edit artifacts within it, he or she can grant viewing access to the collaborator. This becomes possible by clicking on the button labelled “Privacy” (Figure 3.1(d)). Then the user can choose to make their workspace public or private. This helps users to share their findings easily, without the need to move them to the shared workspace. This access is not permanent, so the user can grant and revoke it multiple times to or from the collaborator.

There are only six slots for the charts on each workspace, so if users create more charts, they will need to remove some of them from the workspace. This number of

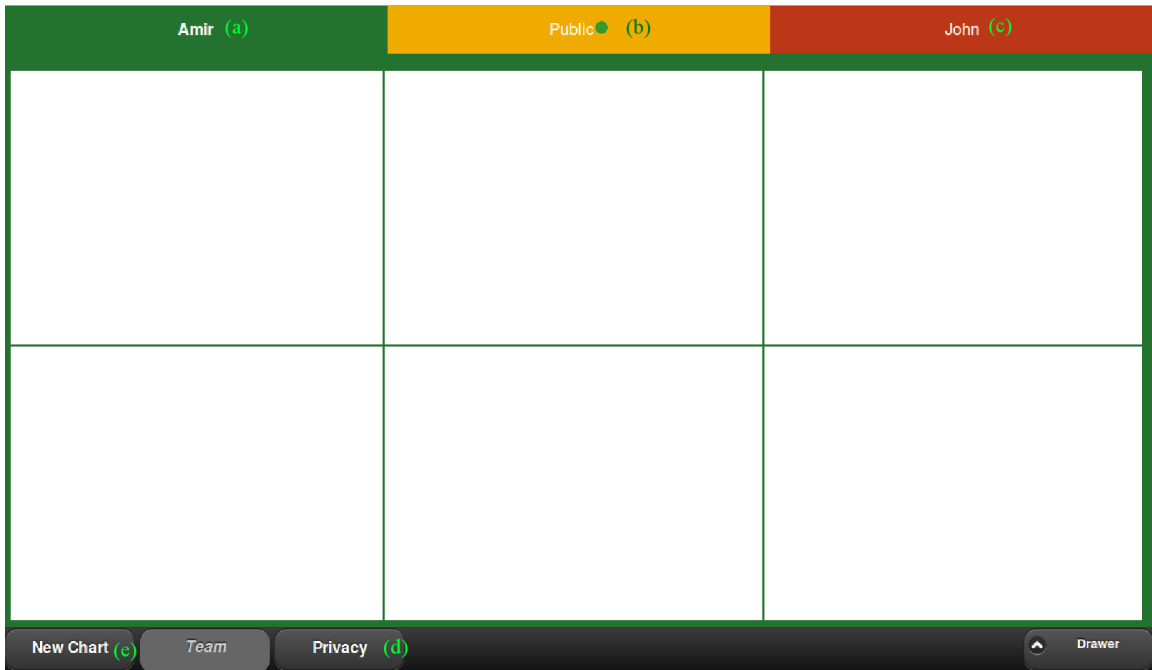


Figure 3.1: Operam workspace view. (a) Name of user who is logged in, (b) tab title for the shared workspace and the green circle representing presence of collaborator in the shared workspace, (c) name of the collaborator, (d) privacy option button, and (e) chart creation button

chart were chosen because six charts could be placed within the screen in a size which was easy to see and users don't need to scroll. However in Operam users do not need to permanently delete the chart. There is a personal drawer provided to each of the users. Users can move any of the charts to their personal drawer. The drawer can be used for temporary storage, but can be also be used to hide some charts from any of the workspaces, as the personal drawer is totally hidden from the collaborators and won't be shared even if users share their workspace with others. So the drawer always remains confidential. One other usage of the drawer is to move the charts between workspaces. A user can move a chart from one workspace to the drawer and then put it back into another workspace.

3.2 Personal Use

As explained earlier, each user will log into their personal workspace by default. To log in, a user enters a name, and clicks on the “Enter the workspace” button from front page (Figure 3.2). This name will be used as the identifier on the top of the



The image shows a login page for a system named "Operam". At the top, the word "Operam" is displayed in a large, white, serif font against a black rectangular background. Below this, the page has a white background. A yellow rectangular box contains the login form. Inside this box, the text "Please enter your name:" is written in a small, black, sans-serif font. Below this text is a white text input field with a light gray border, containing the name "Amir". Below the input field is a wide, white button with rounded corners and a light gray border, containing the text "Enter the workspace" in a bold, black, sans-serif font.

Figure 3.2: Operam login page. Users can enter their name in the field provided.

personal workspace (Figure 3.1(a,c)) and will be shown to the user. In Operam, name is a free-form string for flexibility, e.g. users can log in with their real names, in a collaboration between co-workers, or with their position or company name when collaboration is between two high-level entities.

After entering the system, the user will see a blank screen with six place holders for the charts and a set of options. A user can start by creating a chart. In Operam charts are made based on a sample dataset. User can choose the measures and dimensions and also type of the chart (Figure 3.3(a)) and the chart with those attributes will be created and added to the workspace. The user can always edit a chart by clicking on the “Edit” button (Figure 3.3(c)). Users can also re-order the charts by moving them freely around. However, the charts will always be placed on one of the grid cells.

Having more than one chart, the user can select a data item and see its reflection on other charts (Figure 3.4). This is called brushing and linking. This concept uses the connection between two or more of the visualizations, so that when a data item (e.g. a certain type of product) is selected on one chart, its representation on every other chart will be highlighted (a white border in case of Operam). Assuming that the user has selected a data item on a chart, its reflection on the second chart will follow one of the following cases, depending on the relationship between the data item and the second chart. For this example assume the first chart is showing revenue of different products, and the user has selected a certain product on that chart:

1. If the selected data item is one of the values in the chart’s dimension, representation of that data item will be fully highlighted, e.g. if the second chart is showing number of produced items from same list of products.

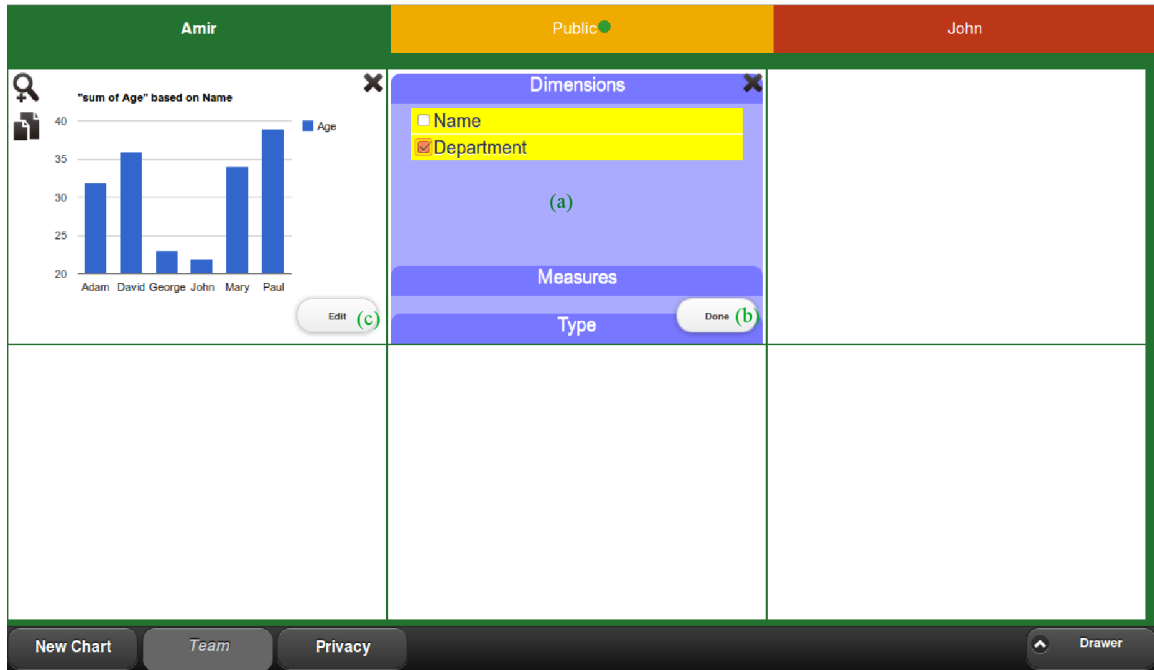


Figure 3.3: Operam chart management. User will select chart dimensions and measures from Dialog for creating a new chart (a), and then press the button labelled “Done” (b). Beside each chart on the workspace, there is a button labelled “Edit” (c) which will switch the chart to edit mode (a)

2. The dimension used in the second chart has a whole-part relation with the data item, but chart dimension is different than that. In this case a portion of measure matching that data item is highlighted. E.g. if the second chart represents revenue of all products in different cities, selecting a certain product on first chart will highlight the portion of revenue in any of the cities gained by selling that certain product.
3. Last, If the dimension of the second chart is completely independent of the data item, we can ignore that chart.

Operam supports first and third items, but later I extended my experiment prototype to include the second option so that it covers all three options. The personal use is though a very small part of the capabilities of Operam. It was developed to support collaboration and so most of its power is in collaborative use of the tool.

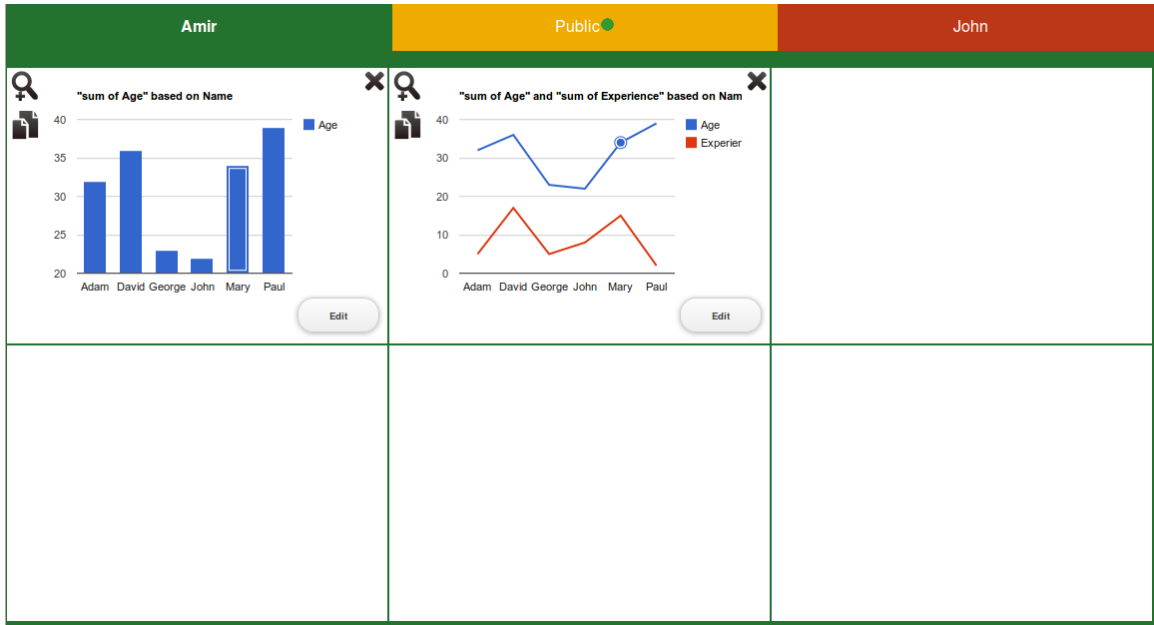


Figure 3.4: Selecting one data item will highlight other representations of it. In this example user has selected “Mary” from first chart (bar with white border), and the value from second chart for the same item (i.e. “Mary”) is highlighted (small circle)

3.3 Collaborative Use

Collaborative use of Operam can start in either of two ways. One person can share their personal workspace with the collaborator using “Privacy” option (Figure 3.1(d)) or alternatively, one can make a shared workspace and ask others to join it. In the latter there is no ownership of the workspace and therefore both collaborators have equal rights to edit and manipulate the workspace.

If someone shares their personal workspace, it will be added as a new tab in the other user’s screen (Figure 3.1(c)). Tabs were used to distinguish the available workspaces and to provide an easy way of switching between them for the users. Users can switch to their collaborator’s tab (which is labelled with his or her name) and see the charts in that workspace. However, a person is not able to edit, move or delete the charts in their collaborator’s workspace. This mode can be used by one of the collaborators to share or demo the findings to other people, while keeping the ownership of the workspace to themselves. A shared workspace will also appear as tab on user’s screen, labelled “Public”.

During a collaborative effort, users will switch between tight and loose collaboration. This is supported by Operam. Users also want to be aware of what the other

person is up to, so they will need to know which tab the other user is on, and what chart they are working with. This type of “awareness” is supported by two annotations on Operam. A user’s presence on a tab is indicated by a green circle beside the tab name, while the interaction with the charts is represented as a colored border around the respective chart.

Operam also supports free-form annotations. User can click on the magnifier icon beside a chart and see the chart in full-screen mode. Then they can draw on the chart using mouse pointer, or by touch (on touch screen devices). The drawings will be shared between users too, and so it can be used to draw the collaborator’s attention to a certain part of the chart.

Another interaction supported by Operam is highlighting. Users can either highlight a chart, or select a data item. In order to highlight a chart, user should click and hold, or tap and hold on a chart. It will hover a semi-transparent layer over the chart for both users. Selecting an item is the same. It will select the data item on any related chart, as well as on both users’ screens. However one limitation of Operam is that only the last item selected will remain highlighted. It is common between collaborator’s too. This approach seemed inadequate, and this led me to ask how I could enable users to highlight items separately, while sharing the selections between them (to enable collaboration) and keeping them aware of each other’s work? In the following chapters, I describe a user study that begins to answer this question.

Chapter 4

Study Design

In my study, I compared three different ways a collaborator’s actions can be visualized: brushing & linking, selection, and persistent selection (see Figure 1.3) in order to assess which provides the greatest awareness while causing minimal interference with individual work. To this aim, I asked participants to perform a primary analysis task, while simultaneously viewing an integrated visualization of what a “remote participant” was doing. Unbeknown to the participants, the remote participant was actually fictitious (implemented as a script, as described in section 4.2). At the end of the study session, I measured participants’ awareness of the remote collaborator’s activities by asking them to identify questions that were in common with the collaborator. I also assessed the visualization interference by measuring completion time of the task, accuracy of their answers, and perceived distraction.

I focused on quantitative measures because I saw this as a useful starting place for investigating feasibility of awareness mechanisms. Another advantage of my approach is that it is not subject to reporting biases; the awareness effects that I hoped to see are not in the focus of the user’s attention during the tasks, and therefore might not be easily self-reported. My hope was that I could identify a mechanism that provided a measurable awareness benefit while not causing substantial interference with individual work. After identifying such a mechanism, then future (perhaps more qualitative) studies could explore nuances of the design and influences on collaborative analysis in a more realistic setting.

The experiment used a between-subjects design with four conditions: a control condition (no visualization of the other person’s actions) plus three visualization conditions (brushing & linking, selection, and persistent selection).

4.1 Participants

I recruited 40 participants among students of the University of Victoria. Participants had various backgrounds such as Engineering, Humanities and Social sciences. They were from different levels of their education: 26 undergraduate students and 14 graduate students. The average age for the participants was 21 years, and there were 20 females and 20 males among them. They were paid \$10 each for their participation. These participants were randomly assigned to one of four different conditions, with 10 participants for each condition. In HCI experiments, typically 10 to 12 participants are assigned to the groups being compared, and it is believed to gain statistical significance for the acquired results.

Before the experiment starts, participants were asked if they have ever worked with data analysis tools. Those who had experience with data analysis or visual analytics tools were then verbally asked about their familiarity with brushing and linking. None of the participants reported being familiar with the concept of brushing and linking.

4.2 Apparatus

Participants were told that a remote collaborator was sharing their screen and working simultaneously, but there was no audio or video link. The remote collaborator was actually simulated and implemented as a script. Prior to the study, I recorded the actions of an experimenter playing the role of a participant. During each study session (except for control conditions), these recorded actions were played back as an animation. Although technically the script was pre-recorded, this setup was designed to emulate synchronous, distributed collaboration. Participants believed that they were seeing the actions of a live collaborator.

This approach reduced the number of participants needed (no need to recruit pairs for each study session). More importantly, the behaviour of the collaborator was identical (i.e., controlled) in each study session and it could not influence the results. The length of the animation was 213 seconds, which covered the whole task time for 22 participants; of the remaining participants, 12 finished their tasks within one minute after the animation stopped.

I used a prototype I developed to enable brushing and linking for visual analysis tasks. Each user worked with a PC running the Ubuntu Linux operating system. The prototype was developed using Javascript language, employing jQuery [1] and D3 [3]

libraries. Participants used Firefox 17 as the web browser to see and interact with the prototype. Screen resolution was 1920 x 1280. Participants interacted with the prototype using a mouse. Time spent on each question was recorded by the prototype.

4.3 Tasks

4.3.1 Primary task (data analysis)

Using my prototype, users could see six charts derived from a tabular data set. The six charts showed one single measure over six different dimensions (refer to Figure 1.1). I used US export statistics from US Census data. In the main phase of the study, the number of contracts was plotted against the following six dimensions using bar charts:

1. Year of export;
2. Quarter of export;
3. Destination of the export (the country);
4. Whether or not the contract was brokered;
5. Decision authority that was responsible for approving or declining the contract;
and
6. Type of the contract (short-, medium-, or long-term).

The participants' primary task was to answer six analytical questions individually using the prototype. Two examples of the questions are "Which decision authority worked on more contracts in decision year 2011?" and "Which countries have more short term contracts than medium term ones?". Users could interact with the system to select one dimension value at a time, and see brushing of their selection on the other five charts (Figure 1.1). All of the questions could be answered using a single selection so multiple selection was not necessary.

4.3.2 Secondary task (awareness)

Participants in the three visualization conditions (conditions other than control) were also asked to perform a secondary task related to the remote collaborator's actions.

These participants were told that they would be seeing some information about the brushing actions of another remote participant who was working on answering similar questions about the same data set using an identical screen. Thus participants would see their own brush actions as well as a visualization of the remote participant's brushing actions. I started by explaining about awareness and the reason it is important in a collaborative setting. I believed it was important for participants to know why they would be presented with a representation of another person's actions, since in real collaborations people would know this. However, I was careful not to draw too much attention to the secondary awareness task, because I was afraid that participants would focus too much of their attention on the collaborator's actions, whereas my design intention was for it to serve as background information, and one of my goals was to determine whether it disrupted primary data analysis work. Instructions explaining the situation were carefully designed, informed by pilot studies, to minimize influence on the performance of the primary task. Thus, apart from this information, participants were not told about the secondary task until the end of the study. At the end, they were informed that three out of six questions they had answered were in common between their task and their collaborator's. Then, they were asked to guess which three questions were in common. This measure was used as the metric for the amount of awareness gained by the user. The remote collaborator's actions were not shown to the control group so these participants were not asked to perform the secondary task.

4.4 Visualization Conditions

I used three different visualization techniques to show the remote collaborator's actions. Each is separately explained below.

4.4.1 Brushing & linking

One obvious way to keep one aware of what a remote person is doing on a shared display is to show an exact copy of their actions on the shared display. My brushing & linking condition did this by visually representing both the remote participant's selected item plus any linked items (as the remote person would see them). To make the visualization distinguishable from the participant's own selection, I used different colors (red for one's own actions, blue for the collaborator's). To make the awareness

visualization less disturbing I used 20% opacity. The color and opacity level were selected after trying various options. I aimed to find a configuration that was clearly visible and distinguishable from the main selection color (red), while also forming a secondary visual layer.

Figure 1.3 (Top right) shows an example of this visualization. The user has selected Quarter 2 from left chart, but simultaneously sees what the other user has selected (i.e. Brazil from right chart) as well as the linked items in other views. This visualization does not explicitly show which bar was selected. This information can be inferred because that bar will be fully colored, but I did not specifically instruct participants about how to do this.

4.4.2 Selection

In contrast to brushing & linking, which blindly mirrors what the remote collaborator sees, we can selectively choose what part is more informative to share among users. In order to understand what a person is interested in, I conjectured that the main selection is the most important part, and the linked items are secondary. Showing the selection only should result in a less cluttered view with more focused information. Therefore, I chose my second visualization to visually represent only the selected item, without any linked items. Figure 1.3 (Bottom left) shows a snapshot of this visualization. In this condition one can precisely see what the collaborator has selected at the present time. The visualization updates when the selection changes or the remote collaborator deselects the currently selected item. The selection is visualized with 20% opacity.

4.4.3 Persistent selection

I assumed that limiting the visualization to the user's selection would make it less cluttered and more focused on important information. However, both the selection and brushing & linking conditions have a transient nature, so the user may miss some of the previous selections. Inspired by [16] and the results from [25], I also designed a persistent selection condition (Figure 1.3, bottom right). In this condition, the user saw current and previous items selected by the collaborator, ordered by their temporal occurrence. Opacity was used to show time since the selection was made. All selections except the current one faded out over time and eventually disappeared, and the current one started disappearing when it was deselected.

Using this visualization, a participant will not only see the most recent selection, but will also have information about earlier selections made by the collaborator. In addition to making it less likely that a selection will be missed, this approach puts the current selection in context. As an example, if the most recent selection is “Brazil”, and we see that the previous selections are also countries, this is an indication of that the person is comparing countries regarding some factor.

In this condition the most recent selection had an opacity of 40%, and each decaying selection had its opacity reduced 1% every second. As a result, each former selection lasted for 40 seconds. During my study, this design meant that the last 4 or 5 selections were usually visible. Also, the questions that were used for the script of the fictitious collaborator were designed such that each required working on a different chart from the previous question; this minimized the overlap of consecutive selections.

4.5 Procedure

Before the start of the study, participants were asked to read and sign a consent form, and fill out a demographic questionnaire. In that questionnaire they were asked about their age, major and level of study, and also if they had experience with visualization applications or had participated in any project or experiment that included data or visual analysis tasks.

After filling in the questionnaire, I presented the prototype to the participant and explained brushing and linking concepts using a sample data set (E-fashion). Then, I worked through two sample questions with the participant to make sure they were comfortable with the concept of brushing and linking. Sample questions were chosen so that they covered different kinds of questions used in the primary task.

When the participants felt confident about brushing and linking, they were briefed about the actual dataset. Different dimensions of the data and their meaning were verbally explained to the participant. At this point, participants in visualization conditions (brushing & linking, selection and persistent selection) were told about their collaborator. The control group participants were informed about the presence of the collaborator, but because they did not see any representation of that person’s actions no further information was provided to them.

Participants were told that the other person would be working on a “different” set of questions, which “may have” some questions in common with them, but would be

randomly ordered. However, I also emphasized that they would be working independently of their collaborator. Here participants in visualization conditions (not control group), were instructed about the visualization assigned to them, as the mechanism of awareness.

At this point they were prepared to start the primary task. Participants had to select each question individually, work with the charts to find the answer and then write down their response on a blank paper provided to them. A drop down list containing six options, labelled as “Question 1” through “Question 6”, was located on the top left corner of the screen (Figure 1.1). Selecting each of the options showed text of the question in the spacing between the drop down and the charts. At the start of the experiment none of the questions was visible to the participant and after making a selection, only one question was present on the screen at a time.

The questions had one to four word answers, where each word was a value in a specific dimension (e.g. “Quarter 1” or “Brazil, China, and India”). Participants clicked on a button labelled “done” (situated on the screen) after they had found the answer but before writing it down on the paper. Timing of the questions was done by the software, and included the time spent to read the question, but not time to write the answer.

After finishing the primary task, participants in visualization conditions were provided with a printed copy of the questions and were told that out of the six questions they had answered, three were in common with their collaborator. They were asked to state which questions they thought were in common by marking beside those questions. I suggested that they could refer to what they remembered from the other person’s actions and see if they could relate the actions to one of the questions.

Finally, participants filled in the post-study questionnaire. The questionnaire was printed on a letter sized paper and handed to the participant. All groups answered a question about ease of the questions for the primary task. There was also an optional field provided for additional comments. The three visualization groups also rated their perception of the distraction caused by the awareness mechanism on a five-level Likert scale.

4.6 Hypotheses

My hypotheses for the study were as follows:

1. All visualization techniques will be better than chance with respect to guessing the common questions (secondary task), but persistent selection will outperform brushing & linking and selection (because it is less transient and therefore less likely to be missed).
2. For completion time on the primary task, the order will be: control < selection and persistent selection < brushing & linking. I anticipate that the three visualization conditions (selection, persistent selection and brushing & linking) may cause some interference with the time to complete tasks and brushing & linking will be the most interfering due to its visual complexity.
3. There will be no differences in the number of errors made in the primary task in visualization groups compared to the control group. I anticipate that people will still complete the tasks, but will just take longer if there is interference.
4. Brushing & linking will be rated as more distracting on the primary task than selection and persistent selection.

Chapter 5

Results

After running the study, I analyzed the data using a number of statistical methods. I first checked whether my measures fit a normal distribution using Q-Q plots. Among the five measures, only completion time matched a normal distribution. Therefore I used ANOVA for analysing time, and the non-parametric Kruskal-Wallis test for other measures.

If a significant effect was identified in any of the measures, pair-wise t-tests (for time) or pair-wise Mann Whitney rank-sum tests (for all other measures) were used to investigate which pairs were significantly different. In all of these cases, I used Bonferroni correction to correct for multiple comparisons.

5.1 Awareness of Collaborator's Actions

I studied awareness of the remote collaborator's activities by counting the number of common questions correctly guessed by the participant (the secondary task). Participants had three questions in common with the remote collaborator, out of the six questions they were working on as their primary task. This measure has a range of 0 to 3, i.e. zero if none of the common questions is guessed correctly, to three if all common questions were correctly identified.

Figure 5.1 illustrates that the number of correctly identified common questions was higher with persistent selection than with brushing & linking or selection conditions. I analyzed the results by comparing the participants' scores to the score predicted by random chance (i.e. 50%). Although participants were not forced to guess all three common questions and in some cases they refused to guess any, correctly guessing the

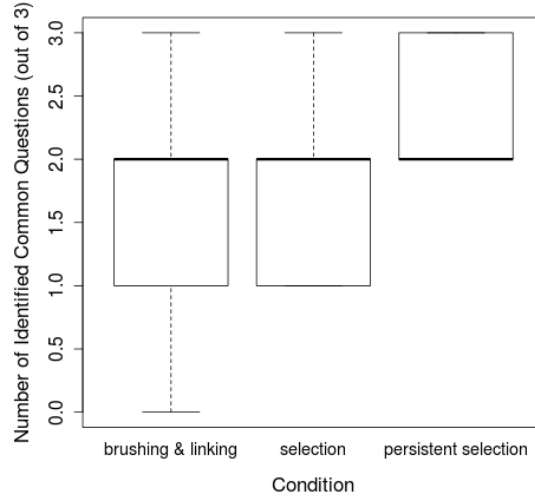


Figure 5.1: Box plots of the number of correctly identified common questions in the secondary task (out of 3).

common questions may have happened by chance. Therefore each of the conditions was compared to a theoretical value of 50% (1.5 out of 3), using a Wilcoxon test. This analysis showed that persistent selection was significantly better than chance ($V=55$, $p < 0.01$), while brushing & linking ($V=29$, $p=0.916$) and selection ($V=41.5$, $p=0.145$) were not significantly better than chance.

I also compared the awareness measures between visualization conditions using the Kruskal-Wallis rank sum test with condition as a between-subject factor. I found a marginally significant effect of condition on awareness ($X^2(2)=5.5$, $p = 0.061$). Running a pair-wise Mann Whitney test with Bonferroni correction showed that persistent selection was more accurate than brushing & linking with marginal significance.

5.2 Primary Task Completion Time

One goal of my study was to determine if adding any of the selected visualizations would result in interference with the user's primary task. I chose time performance as the primary indication of interference. I assumed that if any of the visualizations interfered participants' work, it would result in a longer time to perform the tasks compared to the control condition.

As shown in Figure 5.2, completion time was highest with the brushing & linking

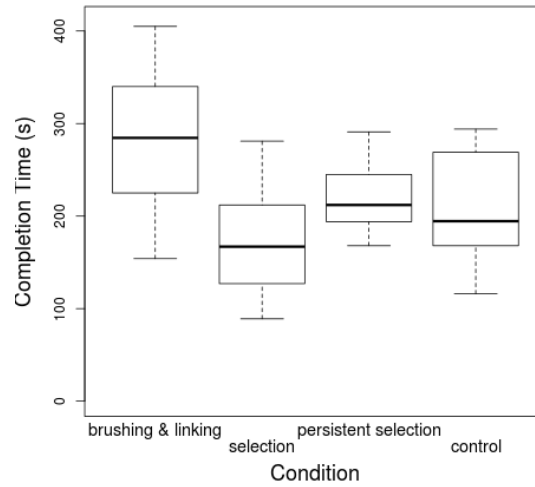


Figure 5.2: Box plots showing primary task completion time for the four conditions.

condition, and lowest with the selection condition. The primary task included six questions, and each took a different amount of time to answer, ranging from 16 to 54 seconds. To account for the effect of question differences on completion time, I ran a two-way ANOVA on time with condition as a between-subjects factor and question as a within-subjects factor. I note that questions were not in random order; however, I do not consider this a problem because comparing between the questions is not the purpose of my analysis.

Results from the ANOVA showed significant main effects of condition ($F(3) = 5.2$, $p < 0.01$) and question ($F(5) = 35.2$, $p < 0.001$), but no significant interaction between question and condition.

After this step, I ran pair-wise t-tests on completion time with condition as the factor, employing Bonferroni correction. The results showed that selection was significantly faster than brushing & linking ($p < 0.01$) and the control group was faster than brushing & linking with marginal significance ($p = 0.051$). Other pairwise comparisons were not significantly different.

5.3 Primary Task Accuracy

Another measure related to interference is the number of correct answers, or the accuracy in answering the questions. Participants had to answer six questions, each having

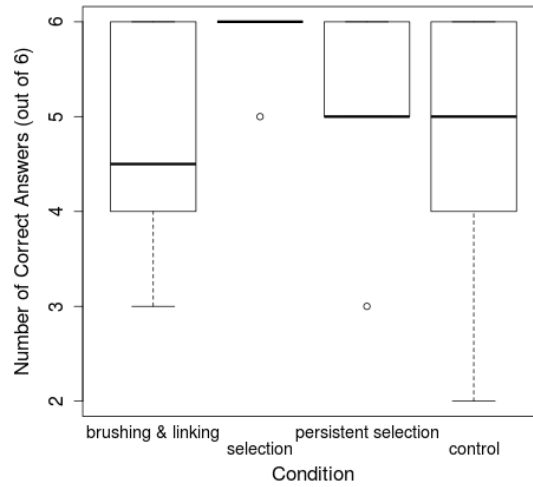


Figure 5.3: Box plots showing primary task accuracy (number of correct answers, out of 6).

exactly one correct answer. Figure 5.3 summarizes the number of correctly answered questions, and illustrates that selection had somewhat higher average accuracy than the other conditions.

The Kruskal-Wallis test showed a significant effect of condition ($X^2(3) = 8.4$, $p = 0.037$). However, post hoc tests were only marginally significant ($p < 0.1$). Corrected pair-wise Mann Whitney tests showed that the selection group was marginally more accurate than the brushing & linking group ($p = 0.098$), and than persistent selection group ($p = 0.051$).

5.4 Perceived Distraction

Results of participants' distraction ratings are shown in Figure 5.4. Brushing & linking tended to be rated as more distracting than the other conditions. The Kruskal-Wallis test showed a marginally significant effect of condition on perceived distraction ($X^2(2) = 5.3$, $p = 0.071$). However corrected pair-wise Mann Whitney tests did not show any significant differences between the three visualization conditions.

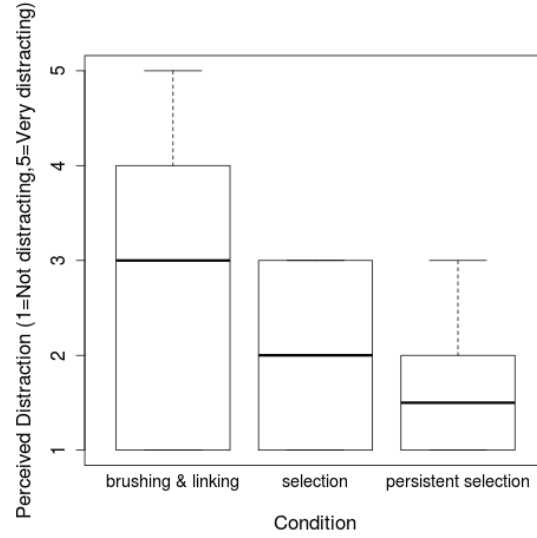


Figure 5.4: Box plots of perceived distraction for the three visualization types. Distraction was measured on a 5-point Likert scale (1=Not distracting, 5=Very distracting).

5.5 Qualitative Comments

During the experiments, I obtained interesting feedback from participants. These comments were usually either mentioned to me after the study, or written in the optional comment field in the post-study questionnaire.

Two participants mentioned that when they saw the other person working, they were encouraged to work faster. They somehow felt an urge to work faster to catch up with their collaborator. This effect was not an intended outcome of my design and was surprising to me. I note that the recorded animation had a duration (213 seconds) that was very close to, but slightly faster than, the average of participants' total time (218 seconds). It was also interesting that those two participants had a total time much lower than the average time for all participants and also for the participants in their condition. One of the participants was from the persistent selection condition and the other one belonged to the selection group.

Also, one participant reported using the awareness mechanism as a way to validate their own approach to the task. The participant stated that they observed the collaborator's method to make sure they were doing the task in the same way. As my participants were unfamiliar with the concept of brushing before participating in this study, this point demonstrates one possible advantage of awareness mechanism, i.e.

assuring the users of their method if it is aligned with others, or correcting them in earlier stages if they are performing it wrong. Of course, this only has the desired effect if the collaborator uses an effective strategy; an inexperienced collaborator could have the opposite effect.

Participants' comments also indicated that my visual encoding was effective at making the awareness information attentionally ambient. A few participants in my brushing & linking condition commented that even though brushing was occupying a big portion of their screen and it was constantly changing, "it really didn't distract" them, "because of its subtle [blue] color and transparency."

Chapter 6

Discussion

In summary, my findings demonstrate that persistent selection enabled the highest awareness of a collaborator's activities, while causing minimal interference with individual work. In this chapter, I first revisit each of my hypotheses to discuss the results in detail. Following this, I discuss more general implications of my results and design considerations.

6.1 Discussion of Hypotheses

H1: All visualization techniques will be better than chance with respect to guessing the common questions (secondary task), but persistent selection will outperform brushing & linking and selection.

This hypothesis was partially supported. Although average scores for both selection and persistent selection groups were higher than chance, my analysis showed only persistent selection to be significantly better than chance. This confirms my hypothesis that persistent selection would provide the best level of awareness. However, it is uncertain whether the other two techniques provided any awareness benefit.

H2: For primary task completion time, the order will be: control < selection & persistent selection < brushing & linking.

This hypothesis was partially supported. I expected all three visualization conditions to be disruptive and therefore the participants in those groups to be slower than the control group. However, my results showed that only brushing & linking was slower than the control group (with marginal significance). This supports my

conjecture that the visual clutter and motion caused by the collaborator’s actions in the brushing & linking condition would interfere with individual work. Although some participants in the brushing & linking condition reported that they did not find the collaborator’s activities disruptive (due to good color choices), this condition nonetheless appeared to cause interference. This might be due to the cluttered nature of the brushing & linking condition [25]. The fact that my other visualizations took no longer than the control is very promising; it suggests that some of the collaborator’s actions can be shared without negatively influencing individual performance. It is worth noting that the average completion time for the selection condition was surprisingly faster than the control group, but this difference was not statistically significant.

Selection and persistent selection were faster than brushing & linking (significant for selection). This demonstrates that simply showing everything that a collaborator sees is not the best solution. A selected representation of a collaborator’s actions can reduce interference while providing the same or better levels of awareness.

H3: There will be no differences in the number of errors in visualization groups compared to the control group (primary task).

Looking at accuracy of answers, I see that selection was better than brushing & linking and persistent selection (with marginal significance). However none of the visualization conditions were significantly different than the control group. Although I cannot statistically conclude that the conditions had identical effects on accuracy (and it is unlikely that they did), my findings are generally in line with my hypothesis that visualization conditions would not be worse than the control group in terms of answer accuracy. Thus, none of the visualizations appeared to significantly interfere with performing the primary task in terms of accuracy.

H4: Brushing & linking will be rated as more distracting than selection and persistent selection.

This hypothesis was not supported. Although brushing & linking had a higher average distraction rating, this effect was not significant. This might be explained by participants’ comments that the color was subtle enough that they could focus on their work and ignore the awareness representations completely.

6.2 Implications for Design

After reviewing hypotheses of the experiment, the research questions can be revisited. First question was that if analysts can gain awareness of their collaborators' activities by viewing a representation of their brushing actions. My results demonstrate that visualizing brushing actions in an attentionally ambient way can enable loosely coupled collaborators to gain awareness of each other's activities within a shared display space.

Second, I was looking to see how visual representation of brushing actions can be improved in order to maximize awareness while minimizing interference with individual work. It turned out that the most effective design for the awareness mechanism is likely to be selective in what is shown (i.e. show only the collaborator's selections, not everything they see) and include some temporal persistence of the awareness visualization (i.e. so that the collaborator's selections do not immediately disappear when the collaborator selects something new).

I found that a cluttered awareness visualization, like the brushing & linking condition, causes interference with a user's primary task. Therefore, selective filtering of the visualization is required. In my study, brushing & linking seemed to be showing more than what was needed for the participant to answer the secondary task. However, what makes an effective summary may be task-specific, and we should take this into consideration when designing attentionally ambient awareness mechanisms for other tasks. We may also need to visualize awareness information with different levels of detail depending on the phase of collaboration and role of the awareness information. Most importantly, I expect that it is important to show more detailed information about brushing & linking during closely coupled work (to support conversations about the data), but to show less detailed information (only selections) when participants work in a loosely coupled fashion.

Persistent selection was the only condition that performed significantly better than chance in informing the participants of their collaborator's activities. This also supports previous findings on peripheral visualization [25] and persistent brushing [16], which suggest that a peripheral visualization must persist to enable it to be noticed. Therefore, in designing peripheral visualization for secondary tasks I could consider slowing down the decay rate of the indicators in order to give more time for the user to notice and switch between their primary and secondary tasks. This may be especially critical when the primary task is attention-consuming or time-constrained.

Finally, my persistent selection visualization, which only shows the selected items, could conceivably be applied to a wider range of actions beyond brushing, as long as the actions involve selecting dimensions or values. It might be beneficial to visually represent the type of action performed, or to represent positive indications of interest in an item (e.g. brushing) differently from indications of non-interest (e.g. filtering).

Brushing, filtering, and other data-centric actions have another benefit: they are tied to the underlying data and therefore can be understood by the machine [13]. Therefore, it may be possible to exploit this attribute to propose intelligent methods to integrate and aggregate users' actions that likely will reduce the amount of data shown to a collaborator.

6.3 Design Details for Further Study

Some design details should be further investigated before directly deploying my visualization approaches to similar collaborative situations. As an example, for my persistent condition, I selected a fall-off speed based on pilot studies with my fictitious collaborator data. The optimal fall-off speed for a broader range of people and situations would likely be different and would need to be determined. Most importantly, the fall-off speed needs to be investigated in the context of a longer, more in-depth data analysis session, since my study sessions were very short. A simple speed-based solution may actually not be optimal either. Developing an intelligent algorithm to identify groups of related actions and set the decay rate based on them may turn out to be a more valuable approach.

Alternative color palettes and transparency levels should be considered too. It should be possible to increase the transparency level, which would reduce the visual salience of the background visualization, while still keeping it detectable. Additionally, my visualization uses the color channel to distinguish one's own brushing and linking actions from those of the collaborator. In visualizations that use color coding for another purpose, this is potentially problematic. For single users, visualizations typically either use saturation to indicate selections, or reserve one color as the "selection color". My awareness visualizations may require reserving two colors instead of one (reducing the space of colors that can be used for color coding) or using three levels of saturation (probably most saturated for one's own selections, second most saturated for the collaborator's selections, and least saturated for non-selected items). Future studies should investigate how these encoding schemes interact with

color coded data.

6.4 Limitations and Future Work

My study should be viewed as a starting point that shows promise for the idea of providing awareness through visualization of brushing actions. However, substantial additional research is needed to explore how to apply this idea in a more diverse range of situations.

To begin with, my study examined single selection (i.e. selecting exactly one bar). This was a logical place to start, but future work should consider the effects of multiple selections. Multiple selections will increase the amount of data shown to a collaborator for selection and especially persistent selection views. The amount of clutter may not be a problem in real practice, or might be mitigated by intelligently aggregating or summarizing the selections, but this issue needs further examination. In this case, I also need to devise ways to visually represent repeated selections in the persistent visualization (e.g. if a person selects Canada & Brazil, followed by Canada & Germany).

I also focused on bar charts, whereas in a more realistic situation analysts use a variety of chart types, such as time series plots and scatter plots. In addition, although my study did involve multiple linked views, analysts are likely to use a larger and more diverse collection of views. I focused on bar charts for simplicity and consistency in the study design. I believe it should be possible to extend the idea to other types of charts, but finding an effective representation of selections will require some design exploration. Since persistent selection was most effective in bar charts, I would like to extend this technique to a variety of chart types and ensure that it can be understood. Multiple linked views (beyond the six in my study) will add an additional level of complexity, and I should investigate how this influences the effectiveness of persistent selection. In practice, I would also expect the two collaborators to be simultaneously examining different views of the same data set (e.g., different charts). This is unlikely to be a big problem for persistent selection as long as the collaborator's selected items can be shown in some currently visible plot. However, when the collaborator's selected items are not included in any of the visible charts, a different approach will be needed. For that case, we may want to highlight the relevant dimension in a visible list of dimensions or add other informative indicators. It will also be useful to explore how to present other actions that may indicate interest in the data, such as filtering

and aggregation.

Evaluating the persistent selection idea in a more realistic context will also require recruiting professional analysts as participants. Since my study focused on basic issues of perceptibility and interference of the awareness visualizations, I was able to use fairly simple tasks. My student participants seemed comfortable with these tasks and were fairly successful at answering the questions. However, more complex tasks would likely be approached differently by professional analysts as compared to students.

My measure of awareness, i.e. number of guessed common questions, showed promising results that people can gain awareness through visualization of each other's brushing actions. I emphasize that even though participants were not informed of the secondary task before the study, those in the persistent selection group were able to use what they recalled, to guess the common questions effectively. Devising a quantitative measure of awareness was actually very challenging. Awareness is a somewhat "fuzzy" concept, and it is not clear what attributes would best indicate that a person is "more" or "less" aware. My common question strategy provides one quantitative metric that I consider useful. However, additional measures of awareness could improve the precision of awareness measurements or add diversity by including other aspects of awareness. For example, it would be useful to have a more holistic way to measure a participant's understanding of their collaborator's activities, or to assess their understanding of a collaborator's higher-level strategy or work process.

The result of this experiment can be applied to Operam prototype (which was the motivation of this study), and other visual analytics tools. One implication of this experiment for Operam is to add temporal persistence to its awareness mechanisms, including the selected item, highlighted chart or the workspace on which the collaborator is currently working.

As my own future work I would like to consider situations where there are more users involved, they work in both loosely and closely coupled styles, and do not share an identical workspace. I would also consider how to apply persistent selection to other types of charts and how to visually represent multiple selections.

6.5 Threats to validity

The validity of this study is subject to a group of threats. In this section I will discuss those threats and my effort in mitigating them. The threats are listed and evaluated under three main categories of threats to internal validity, threats to external validity

and threats for construct validity.

6.5.1 Internal validity

Internal validity is about whether valid conclusions can be drawn based on the internal design of the study. It was concluded that selection is better than brushing & linking because of its selective nature, and persistent selection is good because it adds temporal persistence to the awareness mechanism. There is some possibility that this conclusion may be incorrect. Those two conditions might be better than the brushing & linking condition because of some other differences that were not considered. Therefore some alternative designs should be considered and used in order to examine these assumptions independently.

6.5.2 External validity

External validity relate to generalizability of the results. Despite being careful about the generalizability of the results of a study, there are always some threats to generalizing a result from one situation to the other, or from a sample group to a more-extended group of people.

My study, just like many other HCI experiments, is biased in selection of the participants. The participants of my experiment were among students of University of Victoria, so there is a bias regarding this selection on the results. Though, recruiting participants from different departments of the university will reduce this risk.

On the other hand, the experiments were run in a research lab, and under controlled circumstances. This situation rarely matches the way people work with visual analytics tools. It was also limited to loosely-coupled phase of the collaboration and therefore it may not be generalizable to a real-life collaboration when people have both closely-coupled and loosely-coupled interactions.

To generally overcome the threats to external validity, this study can be performed with different groups of participants and under different circumstances to see if the same result will be acquired. It might introduce threats to internal validity though, as it introduces more parameters which may affect the result in each of the situations and therefore a careful experiment design is required.

6.5.3 Construct validity

Construct validity addresses the relevancy of what has been measured with what the study was supposed to measure. The goal of this study was to find the effect of visualization method on level of awareness and this was evaluated using a measure for awareness. So the validity of the results depends on validity of this measure.

In this study, the number of common questions correctly guessed was used as a measure for awareness. While this measure might be relevant, it is also an indirect conclusion of the amount of awareness gained by the participant. The participant must look at the actions performed by the collaborator and conclude what questions they were working on. It is possible that the actions might relate a similar question and not the exact question user is aware of. I tried to minimize this by designing the questions such that each of them relate to a separate chart, and therefore need different selections for answering them as compared to other questions.

Another example of a threat to construct validity is “hypothesis guessing”. The participants in a study may try to guess what the goal of the study is, and it may affect their way of dealing with the artifact used in the experiment. I tried to prevent it by hiding the actual goal behind the study from the participants. However it is still a threat as having a wrong perception of the study may also cause some unpredicted effects on the results.

In conclusion, I identified and attempted to mitigate various threats to validity. More experiments are required to reduce other types of threats for the validity of the results presented in this thesis, which can be a direction for the future work.

Chapter 7

Conclusions

I compared three methods of visualizing a collaborator's brushing actions on a shared multi-view visualization (brushing & linking, selection, and persistent selection), as a way of providing awareness of the collaborator's activities. I measured which design provides the best level of awareness while maintaining minimal interference with a participant's primary task.

My experiment demonstrates that people can gain awareness of a collaborator's activities by viewing the collaborator's selections as they investigate data through brushing and linking. This is more applicable if the method of awareness is accompanied with a good design. In particular, the persistent selection visualization technique enabled people to perform better than chance on a secondary awareness task, while causing minimal interference with their primary tasks. Other techniques (selection, brushing & linking) were less effective at providing awareness, and brushing & linking caused significant interference. My results reveal that there is value to revealing data selections to collaborators via an attentionally ambient visualization, even when collaborators are working in a loosely coupled fashion. I also demonstrated that only the selection needs to be shown to support awareness, and that persistence of the selection is important to increase the likelihood that it will be noticed.

My results show promise for exploiting analysts' natural brushing actions to build a collaborative awareness mechanism. Future research can build upon these results to design effective awareness mechanisms for collaborative visualization. Other design possibilities can be explored to elevate the level of awareness or to extend it to other situations and different types of analysis.

Appendix A

Study Details

A.1 Questions

A.1.1 Answered by participant

1. Which decision authority worked on almost the same number of contracts during different quarters?
2. Which decision authority worked on more contracts in decision year 2011?
3. What type of contract is mostly authorized by Board?
4. What type of contract doesn't include any brokered deal (i.e. all contracts of that type are not brokered)?
5. Which countries have more short term contracts than medium term ones?
6. In which quarter, not brokered contracts outnumber the brokered ones (i.e. we have a greater number of contracts that are not brokered as compared to contracts that are brokered)?

A.1.2 Answered by fictitious user

1. Which countries have more short term contracts than medium term ones?
2. Which country was not destination of any export in decision year 2010?

3. In which quarter, not brokered contracts outnumber the brokered ones (i.e. we have a greater number of contracts that are not brokered as compared to contracts that are brokered)?
4. Which decision authority works on almost the same number of contracts during different quarters?
5. Which decision authority is mostly responsible for brokered deals?
6. Which countries have same number of contracts between quarter 1 and quarter 2?

A.2 Pre-questionnaire

Department and year of study:

Age:

1- Have you ever worked as a data analyst or any job which included data analysis using graphical charts?

☐ Yes ☐ No

Comment:

2- Have you ever participated in any course project which involved analyzing some data based on charts like bar chart o pie chart?

☐ Yes ☐ No

Comment:

A.3 Post-questionnaire

1- Rate questions based on the following scale:

☐ Very easy ☐ Easy ☐ Normal ☐ Hard ☐ Very hard

Comment:

2- What's your opinion about the following statement:

I was distracted by other participant's work while working on my tasks.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

Comment:

Bibliography

- [1] <http://jquery.com>.
- [2] Lyn Bartram, Colin Ware, and Tom Calvert. Moving icons: detection and distraction. In *Proc. IFIP TC*, volume 13, pages 157–165, 2001.
- [3] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D³ data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.
- [4] AJ Brush, David Barger, Jonathan Grudin, and Anoop Gupta. Notification for shared annotation of digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 89–96. ACM, 2002.
- [5] JJ Cadiz, Gina Danielle Venolia, Gavin Jancke, and Anoop Gupta. Sideshow: Providing peripheral awareness of important information. *Microsoft Research, Collaboration, and Multimedia Group, Tech. Rep*, 2001.
- [6] Yang Chen, Jamal Alsakran, Scott Barlowe, Jing Yang, and Ye Zhao. Supporting effective common ground construction in asynchronous collaborative visual analytics. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 101–110. IEEE, 2011.
- [7] Mei C Chuah and Steven F Roth. Visualizing common ground. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pages 365–372. IEEE, 2003.
- [8] Herbert H Clark and Susan E Brennan. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.

- [9] Craig H Ganoe, Jacob P Somervell, Dennis C Neale, Philip L Isenhour, John M Carroll, Mary Beth Rosson, and D Scott McCrickard. Classroom BRIDGE: using collaborative public and desktop timelines to support activity awareness. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 21–30. ACM, 2003.
- [10] Darren Gergle, Robert E Kraut, and Susan R Fussell. Language efficiency and visual technology minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, 23(4):491–517, 2004.
- [11] Tom Gross, Wido Wirsam, and Wolfgang Graether. Awarenessmaps: visualizing awareness in shared workspaces. In *CHI’03 Extended Abstracts on Human Factors in Computing Systems*, pages 784–785. ACM, 2003.
- [12] Jonathan Grudin. Why cscw applications fail: problems in the design and evaluation of organizational interfaces. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, pages 85–93. ACM, 1988.
- [13] Jeffrey Heer and Maneesh Agrawala. Design considerations for collaborative visual analytics. *Information visualization*, 7(1):49–62, 2008.
- [14] Jeffrey Heer, Fernanda B Viégas, and Martin Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1):87–97, 2009.
- [15] Petra Isenberg and Danyel Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. In *Computer Graphics Forum*, volume 28, pages 1031–1038. Wiley Online Library, 2009.
- [16] Petra Isenberg, Danyel Fisher, M Ringel Morris, Kori Inkpen, and Mary Czerwinski. An exploratory study of co-located collaborative visual analytics around a tabletop display. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 179–186. IEEE, 2010.
- [17] Ruogu Kang and Sara Kiesler. Do collaborators’ annotations help or hurt asynchronous analysis. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, pages 123–126. ACM, 2012.

- [18] Qing Li and Chris North. Empirical comparison of dynamic query sliders and brushing histograms. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 147–153. IEEE, 2003.
- [19] Narges Mahyar, Ali Sarvghad, Melanie Tory, and Tyler Weeres. Observations of record-keeping in co-located collaborative analysis. *HICSS 2013*, Jan 2013.
- [20] Gloria Mark and Alfred Kobsa. The effects of collaboration and system transparency on CIVE usage: an empirical study and model. *Presence: Teleoperators & Virtual Environments*, 14(1):60–80, 2005.
- [21] D Scott McCrickard, Richard Catrambone, and John T Stasko. Evaluating animation in the periphery as a mechanism for maintaining awareness. In *Proceedings of the IFIP TC. 13 Conference on Human Computer Interaction*, pages 148–156, 2001.
- [22] D Scott McCrickard and CM Chewar. Attuning notification design to user goals and attention costs. *Communications of the ACM*, 46(3):67–72, 2003.
- [23] Zachary Pousman and John Stasko. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the working conference on Advanced visual interfaces*, pages 67–74. ACM, 2006.
- [24] Zachary Pousman, John T Stasko, and Michael Mateas. Casual information visualization: depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1145–1152, 2007.
- [25] Jacob Somervell, D Scott McCrickard, Chris North, and Maulik Shukla. An evaluation of information visualization in attention-limited environments. In *Proceedings of the symposium on Data Visualisation*, volume 2002. Citeseer, 2002.
- [26] James Tam and Saul Greenberg. A framework for asynchronous change awareness in collaborative documents and workspaces. *International Journal of Human-Computer Studies*, 64(7):583–598, 2006.
- [27] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1181–1190. ACM, 2006.

- [28] Chris Weaver. Is coordination a means to collaboration? In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*, pages 80–84. IEEE, 2007.