

How can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation

Youn-ah Kang, Carsten Görg, *Member, IEEE* and John Stasko, *Member, IEEE*

Abstract—Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative evaluations. Determining how such systems foster insight and sensemaking is important for their continued growth and study, however. Furthermore, studies that identify how people use such systems and why they benefit (or not) can help inform the design of new systems in this area. We conducted an evaluation of the visual analytics system Jigsaw employed in a small investigative sensemaking exercise, and we compared its use to three other more traditional methods of analysis. Sixteen participants performed a simulated intelligence analysis task under one of the four conditions. Experimental results suggest that Jigsaw assisted participants to analyze the data and identify an embedded threat. We describe different analysis strategies used by study participants and how computational support (or the lack thereof) influenced the strategies. We then illustrate several characteristics of the sensemaking process identified in the study and provide design implications for investigative analysis tools based thereon. We conclude with recommendations on metrics and techniques for evaluating visual analytics systems for investigative analysis.

Index Terms—Visual analytics, information visualization, evaluation, investigative analysis, user study.

1 INTRODUCTION

Recent years have seen a rise in the number of visual analytics systems built to assist investigative analysis (IA). Many stimuli are behind the development of these systems including the availability of example data sets via contests and challenges [1], the increasing importance of this type of work to government and intelligence activities [2], and the emergence of the visual analytics area in general.

Although many new investigative analysis systems are being built, we still do not well understand how to evaluate and assess them. Evaluating interactive visualization systems is challenging in general [3], but investigative analysis scenarios add even more difficulty. Going beyond the typical goals of information visualization such as identifying correlations, outliers, etc., investigative analysts perform sensemaking activities, develop hypotheses about the data, and seek to understand it more thoroughly. One often thinks of analysts “connecting the dots” or “putting the pieces together.” Ultimately, analysts seek to develop insight about the data, a challenging activity to identify and measure [4].

One area in particular lacking much research is the controlled, comparative evaluation of investigative analysis systems. A number of systems have been studied in trial usage scenarios by trained analysts [5], [6], [7], but these studies did not compare performance against other systems or more traditional, “low-tech” approaches.

In this study, we examined use of the Jigsaw system [8] in an analysis scenario as compared to three other investigative methods including paper-and-pencil and simple desktop electronic document storage and search. While we were curious if Jigsaw would prove beneficial, the purpose of our study was not to evaluate Jigsaw per se. Instead, our primary goal was to better understand how visualization can assist investigative analysis, if it truly can. We wanted to see how people would approach data analysis using a visual analytics system. What characteristics of the system, if any, lead to the main benefits? We believe that a comparative scenario where one can examine people working on the same problem under different conditions, although limited in certain ways, does provide a valuable context to address these questions.

A second goal of this research was to better understand evaluation methodologies for investigative analysis systems in general. What should evaluators count, measure, and observe in order to determine the utility of systems? Identifying metrics for visual analytics system evaluation is challenging [9] and is important to organizations making decisions about which systems, if any, to use in practice.

This study is one of the first comparative experiments conducted in this area. We evaluated four settings for analysis with one of these using Jigsaw. 16 study participants performed an investigation in one of the settings. Each participant was given the same data collection containing 50 plain text documents each about a paragraph long. The documents simulated intelligence case reports and participants needed to identify an embedded terrorist plot within the allotted 90 minutes.

This article is an expanded version of the conference paper [10] and it provides more depth and analysis of participants’ performance by setting and by investigative strategy

• Y. Kang, C. Görg, and J. Stasko are with the School of Interactive Computing & GVU Center, Georgia Institute of Technology, Atlanta, GA, 30332.
E-mail: ykang3@gatech.edu, {goerg, stasko}@cc.gatech.edu

employed. We also expand the discussion section, include more feedback from participants, and compare our results to those found in another study [11] examining professional analysts' processes and performance.

2 RELATED WORK

Few experiments have investigated the utility of visual analytic tools for investigative analysis. A study by Bier et al. [5] assessed the suitability of their Entity Workspace System in the context of design guidelines for collaborative intelligence analysis. The researchers modified their system based on five design guidelines and evaluated the system in both a laboratory study with intelligence analysts and a field study with an analysis team. Relying on analysts' subjective feedback in conjunction with quantitative logging data, they confirmed the positive effects of the tool on collaboration and the usefulness of the design guidelines for collaborative analysis.

Perer and Shneiderman [12] recognized the limitations of traditional controlled experiments in examining the process of exploratory data analysis and developed an evaluation methodology for studying the effectiveness of their system, SocialAction. Consisting of a long-term case study [13] and in-depth interviews, the evaluation confirmed the core value of SocialAction - integrating statistics with visualization - and further provided guidance for redesign of the tool.

Several studies have captured and characterized the work practices and analytical processes of individual or collaborative analysis through a qualitative approach. Pirulli and Card [14] studied analysts and developed a notional model of the analytic processes they follow. Chin et al. [11] conducted an observational case study with professional analysts in which participants worked on real-world scenarios, either as an individual analyst or as an investigative team. The researchers revealed various characteristics of the analytical processes of intelligence analysts, such as the investigative methodologies they apply, how they collect and triage information, and how they identify patterns and trends.

Robinson [15] examined how analysts synthesize visual analytic results by studying domain experts conducting a simulated synthesis task using analytical artifacts printed on cards on a large paper-covered workspace. Based on analysis of video coding results, he identified several characteristics in the process of synthesis such as the use of different approaches to collaborative synthesis, a variety of organizational metaphors when structuring information, and the importance of establishing common ground and role assignment.

While these latter three studies did not evaluate specific visual analytic tools or features per se, they provide valuable implications to inform design directions for future support tools. Scholtz [9] emphasizes that the development of metrics and methodologies for evaluation is necessary to help researchers measure the progress of their work and understand the impact on users. She argues that the evaluation of visual analytic environments requires researchers to go beyond performance evaluations and usability evaluations, and proposes five key areas to be considered as metrics and methodologies for evaluation: situation awareness, collaboration, interaction, creativity, and utility.

3 STUDY DESIGN

We recruited 16 graduate students (8 female) from Georgia Tech to participate in the experiment. We explicitly described the study goals and their simulated actions as an intelligence analyst to find students who would be interested and motivated by such a scenario. Participants received either a \$30 or \$40 gift card, depending on their setting and experiment duration, as compensation.

3.1 Task and Dataset

We told participants that they would be taking on the role of a government intelligence analyst. We gave them 50 documents, described as intelligence reports, and asked the participants to identify a hidden terrorist plot.

For this task, we adapted documents from an exercise we had learned about from a military intelligence college. Embedded across some of the documents are hints to a fictional terrorist plot with four sub-stories that support the plot. The main plot is an attack on U.S. airports with surface-to-air missiles, and the sub-stories involve the acquisition and movement of the weapons to the pertinent locations. Each document was a few sentences long. 23 of the documents contained information useful to identifying the threat. The other 27 documents described other suspicious activities but were not relevant to the main plot. Below are three of the documents from the experiment.

Report 5: February 21, 2002

Report of theft of small arms ammunition, four M-16 rifles, six grenade launchers and six surface-to-air missiles. This theft occurred at a military installation in Colorado. [This theft is still under investigation.]

Report 15: August 28, 2002

On August 23, 2002 customs officers seized 234 pounds of cocaine from a marina south of Tampa. Arrested were Jesus Vazquez and Robert D'Onfrio. Each has connections to a South American syndicate and drug cartel.

Report 16: September 7, 2002

Officials seized documents at 134 Boynton Beach Lane in Miami on September 5, 2002. The initials C.G., Y.K., and J.S. appeared along with lists of 12 U.S. airports and various dates in October and November.

We told participants that they needed to identify the plot and ultimately write a short narrative describing the potential threat. In addition, we gave participants task sheets adapted from the VAST Symposium Contest [1], which contained tables for them to list key players, events, and locations relevant to the plot.

3.2 Settings and Procedures

We created four settings in the experiment and assigned each participant to one of the conditions. Each setting had both male and female participants. In setting 1 (Paper), we gave participants the reports as paper documents and asked them to perform the task without any technological aid. In

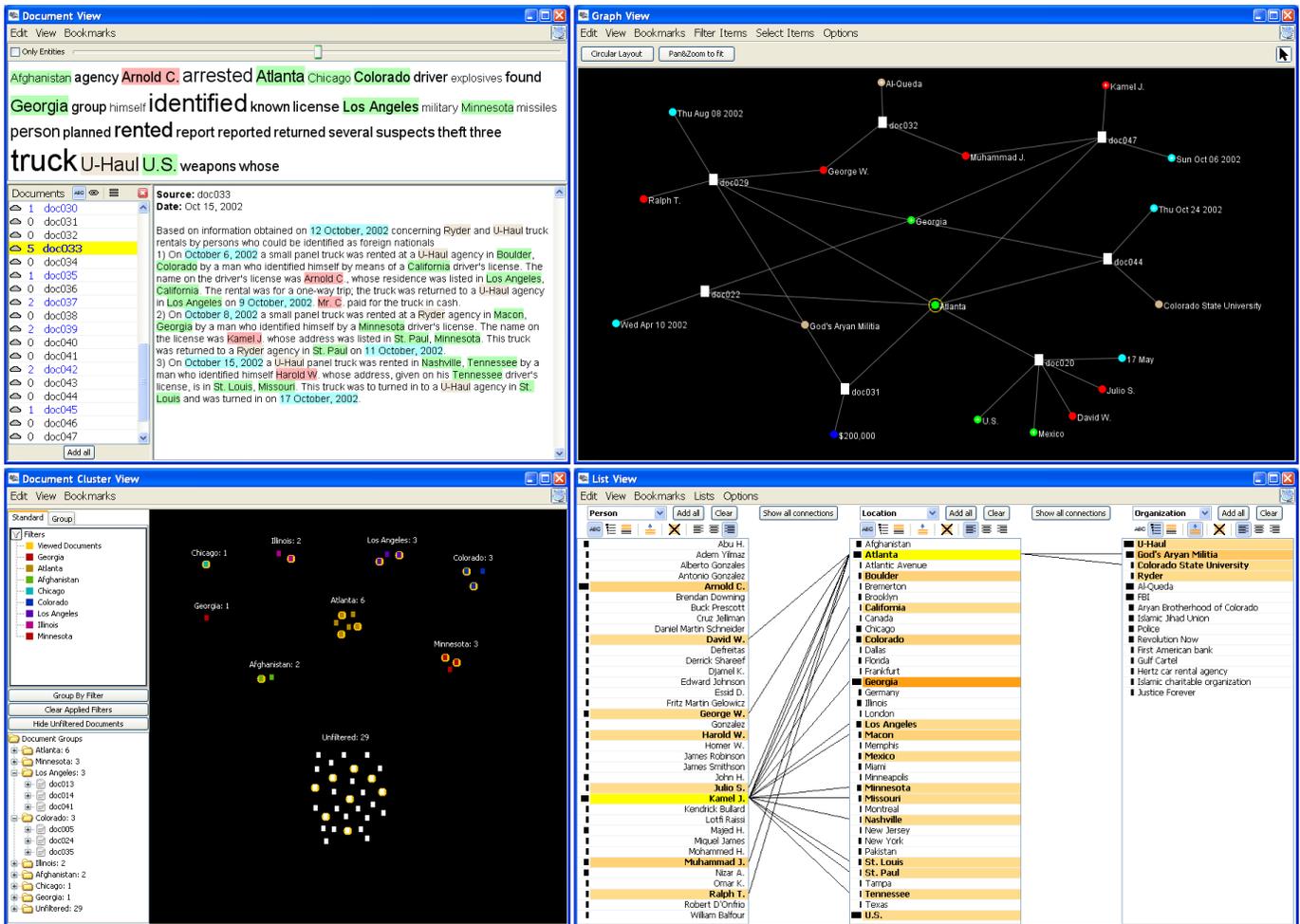


Fig. 1. Jigsaw's Document View, Graph View, Document Cluster View, and List View.

setting 2 (Desktop), we gave participants the documents as separate text files on a computer and made Microsoft Desktop Search available to search for keyword(s) in the documents. In setting 3 (Entity), participants used a limited version of Jigsaw, in which only a modified version of the Document View (tag cloud removed) and text search capability were available. Essentially, this setting was like Desktop except that the Jigsaw Document View highlights identified entities such as people, places, and organizations in the documents. In setting 4 (Jigsaw), participants performed the task using the Jigsaw system. We provided participants in this setting with a short training video of the system three days before the session and gave them an additional 30 minutes of training at the beginning of the session. Neither of these training sessions involved information related to the task used for the evaluation.

In all settings, participants could take notes using pen and paper. We collected the notes and used them to help in our analysis later. For the Desktop, Entity, and Jigsaw settings, participants worked on a four-monitor computer. We gave each participant 90 minutes to work on the problem and conducted a semi-structured interview after each session. In the interview, we asked several questions about their experience including: (1) How they performed the task (i.e., what was their strategy), (2) What was the most difficult aspect, (3) What helped them

perform the task, (4) When were they sure that they were progressing towards goals, and (5) How they would change their approach and process if they were asked to do it again. We also video-taped all the sessions.

3.3 Jigsaw

Jigsaw is a system for helping analysts with the kinds of investigative scenarios encountered in this study. It is a multi-view system, including a number of different visualizations of the documents in the collection and the entities (people, places, organizations, etc.) within those documents. Figure 1 shows some of the visualizations: the Document View (top left) displays documents and highlights identified entities within them, the Graph View (top right) shows connections between documents and entities using a node link diagram, the List View (bottom right) shows connections between entities that are arranged in lists accordingly to their type, and the Document Cluster View (bottom left) displays the documents in the collection arranged by keywords.

A textual search query interface allows users to find particular entities and the documents in which they occur. In addition, entities and documents can be explored directly by interacting with those objects in the views. For instance, new entities can be displayed and explored by user interface operations in

TABLE 1
Study results and statistics, grouped by setting. The measures are explained in Section 4 and Section 5.

	Paper				Desktop				Entity				Jigsaw			
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Grading Task Sheet	-1.75	17	4	13	13	10.5	10.5	-3.5	5.5	-8.25	4	7.5	14.5	13.5	7	17
Grading Debriefing	2	2.5	1	5.5	3	4	1.5	3	3.5	2.5	1.5	6	6	2.5	5.5	5
Final Score	22.87	65.00	24.26	87.08	62.08	67.13	42.13	29.41	52.23	15.00	29.26	81.19	95.05	58.07	75.20	90.00
Performance	fair	very good	fair	excellent	very good	very good	good	fair	good	poor	fair	excellent	excellent	good	very good	excellent
Avg. Score/Setting	49.80				50.19				44.42				79.59			
Documents Viewed	50	50	50	50	50	50	50	50	49	31	45	50	31	50	46	23
Number of Queries					19	18	48	8	23	61	59	91	44	4	26	8
First Query					40:49	19:55	2:47	12:41	1:31	0:29	0:59	3:12	0:18	5:35	25:37	4:18
Amount of Notes	many	none	many	some	many	some	few	some	some	none	none	few	some	few	few	few
First Note Taking	0:07	–	0:05	0:16	1:53	19:57	2:47	8:20	2:37	–	–	3:14	0:48	0:32	5:15	78:48
First Task Sheet	43:20	32:53	70:13	3:25	61:35	20:26	7:33	64:11	28:09	0:52	2:55	7:20	48:26	41:48	43:00	5:33
Strategy Used	OFD	OFD	BFD	OFD	OFD	OFD	FCFT	BFD	BFD	HTK	HTK	FCFT	FCFT	HTK	OFD	FCFT

the views that expand the context of entities and documents. In practice these two approaches are often combined: search queries serve to jump-start an exploration and view interaction then yields richer representations and exploration. The Jigsaw system is described in detail in [8].

3.4 Performance Measures

We created a solution to the exercise and described it in a short text narrative. In addition, we completed the task sheets (relevant people, events, places). Two external raters used this material to grade the anonymized task sheets and debriefings.

For the task sheets the raters awarded each correct item 1 point while misidentified items (false positives) lost 1 point. This grading rule yielded a few negative scores for participants who listed more false positives than correct answers. The maximum reachable score was 29 points. The raters also subjectively graded each narrative debriefing on a scale from 1 to 7, where 7 indicates “Highly accurate; Hits the main plot; Covers all of the supporting evidence and sub-stories” and 1 indicates “Fails to find the main plot; No relevant sub-stories; Points out irrelevant facts and events.” We averaged the scores from two raters for final scores.

3.5 Data Collection and Analysis

Throughout the study, we primarily focused on collecting qualitative data such as observations, follow-up interview notes, and video recordings of all sessions. Wherever possible, we also collected quantifiable data such as the number of documents viewed and the number of queries performed. In order to more closely look at usage patterns of Jigsaw, we logged user interactions and view operations.

During the early, exploratory phase of analysis, we used an inductive approach to examine qualitative data. As we read through notes from observations and interviews, potential concepts and categories emerged, including participants’ strategies, Jigsaw’s influence to the analysis process, and characteristics of sensemaking. These broad categories were refined with specific incidents, anecdotes, and examples from detailed analysis. Detailed info such as video and log data were used as supplements to observation notes when we wanted to further examine and clarify findings. We scrutinized the videos and the log visualization after we identified investigative strategies, in order to verify each participant’s process. As the

analysis evolved, inductive and deductive analysis were used concurrently because some concepts developed further than others.

4 RESULTS AND ANALYSIS

The first block of rows in Table 1 summarizes the performance results of the participants by setting. We normalized the ratings from the task sheets and the debriefing (equally weighted) to a 100-point scale to determine a final score and grouped them into five categories (poor, fair, good, very good, excellent) via quintile rankings.

Our focus here was not on producing statistically significant differences. With such a small subject population, it seems doubtful that such results could even be found. Instead, we view these results as suggestions of overall performance and we relate them to more qualitative findings discussed later.

Within that context observe that participants in the Jigsaw setting earned excellent, excellent, very good and good ratings. If we average the final scores of the four participants in each setting, those using Jigsaw clearly outdistanced those in the other three settings that produced similar average final scores. P4 (Paper setting) and P12 (Entity setting) also performed excellently.

4.1 Activity Patterns

Because of the explorative nature of the task, we were curious about general activity patterns such as how many of the documents were viewed in total, which document was viewed most, and how many times each document was viewed. We also determined how many search queries a participant performed and when the first query was performed. For those participants who took notes on paper, we identified when they first started note-taking, as well as how many and what kind of notes they took. Additionally, we identified when each participant first began completing the task sheets.

Ten of the sixteen participants viewed all the documents at least once (second block of rows in Table 1). Curiously, all of the Paper and Desktop participants read all of the documents, but only one in each of the Entity and Jigsaw settings did so.

The frequency of search queries issued by each participant varies, ranging from 4 times to 91 times (third block of rows in Table 1). (Obviously, participants in the Paper setting could not issue queries.) Overall, those in the Entity setting issued

more queries and start doing so relatively early in the session. Large individual differences existed in all settings, depending on how much each person relied on queries in their analysis.

The fourth block of rows in Table 1 summarizes participants' note-taking and task sheet completion behavior. Thirteen out of 16 people took notes on paper, and those in the Paper and Desktop settings took relatively more notes. Participants mostly jotted down important names, places, and events along with the document number. Some drew a simplified map and used arrows to illustrate traces of people. Most participants started taking notes quite early in the process. In particular, those in the Paper setting typically began taking notes as soon as they started reading. The time at which each participant began to complete the task sheets varied; some people worked on them right after figuring out certain pieces of information such as repeated names, locations, or events relevant to the plot. Most read the documents and concurrently worked on the task sheets. Several participants—P1, P2, P6, P9, P13, P14, and P15—started the task sheets in the middle of the process, when they had confidence about their hypothesis to some degree. P3 and P8 waited to complete the task sheets almost until the end of the session, and it turned out that they still had not determined what to write.

4.2 General Patterns by Setting

In addition to general activity patterns of the participants, we examined how participants in each setting approached the analysis.

Paper Setting

Participants in the Paper setting showed unique activity patterns in part because it was the only setting that did not include any technological aid. Since they were given paper documents, participants had to classify documents physically. Two participants laid out all the documents on a big table, organized by month or by connectedness (e.g., people of the same organization). Two held all documents in their hands and classified them into important or unimportant groups. They placed documents that they considered unimportant on the table face-down. When they needed to refer to one of these documents again, the participants turned them over again. People in this setting mostly made annotations directly on the documents including related document numbers, persons, and events. Most took notes on separate pieces of paper and logged important information from each document, making it easier to look up again. They also made connections between the documents, names, locations, and events by drawing lines and arrows on the notes. Obviously, participants in this setting largely relied on reading for the analysis and each one read every document in the collection.

Desktop Setting

Participants in this setting received 50 separate text files and could use Microsoft desktop search for keyword search. Each participant opened and read each text file one at a time except for one participant who merged all the text files into a single Microsoft Word document. Even though search was available, participants seemed to use it as a supplementary tool and

mostly focused on reading all the documents. On average, each person issued 23 search queries, ranging from 8 to 48. Some queries were repeated; participants often searched for the same keyword multiple times. Interestingly, two participants in the Desktop setting arranged the text documents according to their own classification scheme and put effort into managing documents in the space available. The participants had four monitors to use, so they grouped documents, for example, by locations, and placed each group on a different monitor. One participant resized each document and positioned them on screens so that she could see as many documents as possible. Because these participants did not want to clutter the screen, they often closed and reopened the files as they performed analysis. All participants in this setting took notes on paper.

Entity Setting

The Entity setting was similar to the Desktop setting except that entities were highlighted within the documents. Participants were given the Document View of Jigsaw that showed each entity with a faint background color unique to the entity type (e.g., people, place, organization). Participants also could search for keywords, including entities, using Jigsaw's search capability.

To better understand how Jigsaw (or a portion of it as in this setting) was used, we implemented a logging mechanism and recorded selected user interactions and view operations including search queries, the display of documents in the Document View, and all other view actions (in the Jigsaw setting). Since the display or bringing forward of a document in the Document View does not necessarily mean that the participant actually read the document, we decided to impose a criterion on this measure: we consider a document as being read if it was displayed in a Document View for at least five seconds.

Figure 2, at the top, shows an overview of the usage pattern of the different views for the eight participants in the Entity and Jigsaw settings. Each row of pixels in the maps represents one minute and the color encodes the view being used (active window) by the participant at that time. Gray shows periods when participants worked on the task sheets (no active Jigsaw window). With respect to the processes of participants in the Entity setting, the maps for P10, P11, and P12 are relatively consistent; the map for P9 is slightly different since it has longer note taking periods.

The entity highlighting from Jigsaw seemed to help participants scan the documents faster, focusing more on essential information such as names, places, and events. While participants in the Desktop setting used search more in a supplemental fashion, participants in the Entity setting used search more often, employing it as a guide for which documents to read next. They issued 58.5 search queries on average, ranging from 23 to 91 queries. All four participants issued their first query very quickly after starting the session. Search terms were often repeated as in the Desktop setting. Three participants who totally relied on search-based exploration did not read all the documents, instead focusing only on those brought up by search. People in this setting did not take many notes; two did not even take any notes.

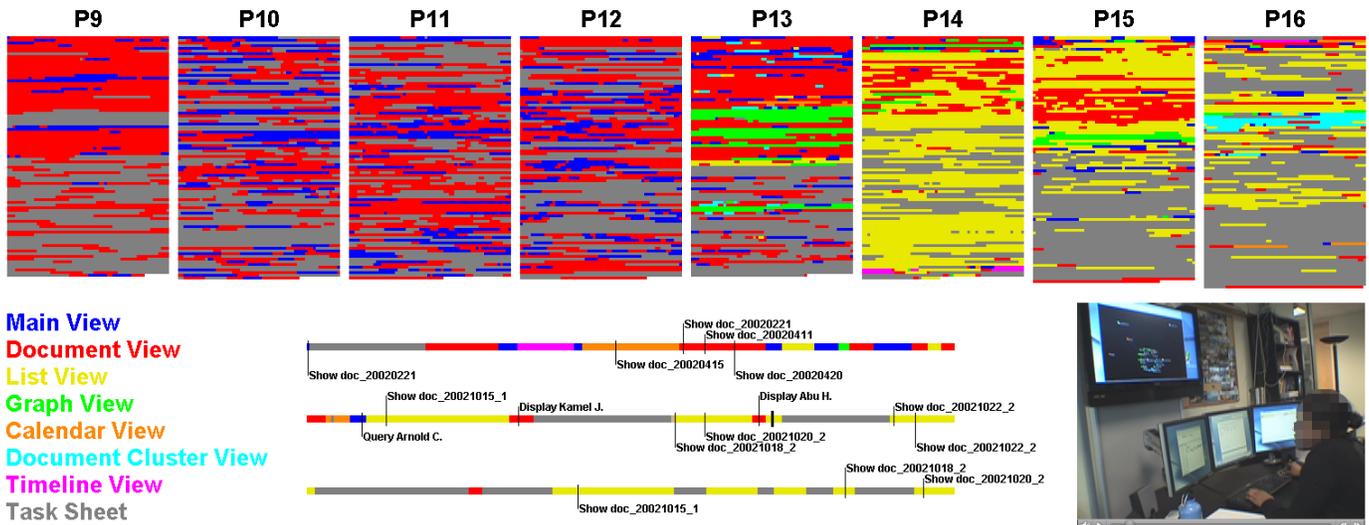


Fig. 2. Overview of the Jigsaw usage patterns (at the top) and an extract from the detailed usage pattern with video for P16 (at the bottom). P9-P12 were in the Entity setting, so they accessed only the Main View and the Document View in Jigsaw. P13-P16 used the full system.

Jigsaw Setting

Participants using Jigsaw all began analysis by using the Document View. They subsequently opened further views such as the List View, the Graph View, and the Document Cluster View. Though they tried different views in the beginning of the analysis, these participants tended to later stick with one more view in addition to the Document View. The most frequently used view was the List View. Three participants used it and the Document View for the majority of the session. They sometimes added the Graph View and the Document Cluster View, but typically just for additional exploration rather than for major investigation. These three participants took a people-centered exploration approach; they put people and documents in two columns in the List View and selected names to read associated documents in the Document View. Conversely, P13 rarely used the List View but instead often used search to find documents of interest. Participants in the Jigsaw setting issued 20.5 search queries on average, ranging from 4 to 44. While reading documents was an important activity, three participants in this setting did not read all the documents. Instead, they selectively read documents using the List View and search function. They rarely took notes, but when they did, the notes were about hypotheses or confirmed facts.

Figure 2 shows the specific activity maps for the participants in the Jigsaw setting. P13 worked primarily with the Document and the Graph View (the List, Document Cluster, Calendar View were also used); P14 primarily used the List View (the Document, Timeline, and Graph View were also used); P15 focused on the List and Document View (the Graph View also was used); P16 used all the views and focused on the List and Document Cluster View. The bottom of the figure shows a small portion of the detailed usage pattern for P16. Each pixel row represents four minutes and the colors again encode the active views. The rows are annotated with queries issued and documents viewed. This view was synchronized with the recorded video (bottom right) using a slider, and we used it to help our analysis of the sessions.

5 INVESTIGATIVE STRATEGIES

Although a participant’s setting clearly influenced how they performed, another important factor emerged to us after repeatedly viewing the video and log data for the 16 sessions. We noticed the participants employing different styles of investigative strategies in their analyses. While the boundaries between the participants’ methods were not always clearly defined, we noticed four general investigative strategies being used by the participants, independent of the setting they were in. In this section, we explain each of these four strategies and describe one or two participants’ analysis process for each in more detail.

5.1 Strategy 1: Overview, Filter, and Detail (OFD)

The most commonly used strategy was “Overview first, filter and select, and elaborate on details,” a strategy quite similar to Shneiderman’s InfoVis mantra [16]. Six participants out of 16 performed analysis using this strategy (fifth block of rows in Table 1). They began by quickly scanning all documents and building up initial ideas of the plot. While gaining an overview, most people jotted down important keywords with corresponding document numbers, drew circles and lines to indicate connections between keywords and documents, and later used these notes as an index for finding relevant documents. After scanning all documents, they revisited relevant documents selectively - either by directly looking up the document or by searching for a keyword that stood out. Then they read each one carefully, extracting key information for the task sheets.

We speculate that this strategy worked well in this task because the dataset was relatively small. Participants were able to gain an initial idea of the important documents or keywords by simply scanning all documents. However, because they made a decision about the importance of each document or keyword based on their own subjective judgment, sometimes they missed important details.

Example: P2 (Paper setting)

Participant 2 began the analysis by reading documents one by one. After reading 10 documents, she tried to cluster the documents by similarity of content. However, this classification did not work well and she started the process again. This time, she scanned every document and grouped documents into those that seemed relevant to the plot and those that seemed irrelevant. The first document she read was about a theft of arms ammunition in Atlanta and she used it as the criterion for being relevant. After scanning through all the documents, P2 began reading the first collection of relevant documents very carefully. At this point, she started filling out the solution templates. In the post-session interview, she said that she had focused on “Colorado” and “theft” and filled out anything relevant to them. After reviewing all the documents in the seemingly relevant group, she turned to the set of irrelevant documents and went through them quickly. Here, she was scanning for potentially relevant documents that she had missed previously. After identifying a few, she read these documents carefully and filled out templates. Next, she worked on the summary debriefing and finished the analysis.

5.2 Strategy 2: Build from Detail (BFD)

The strategy, “Build from detail”, contrasts the previous one. Three participants used this strategy. They started the analysis from details of each document by carefully reading it. Even though they used the search function when important phrases or words arose (where applicable), it was more of an auxiliary use than a main focus. They issued relatively few queries. Instead, they focused on every sentence of the documents, in the fear of missing any relevant information. Some tried to write down important keywords for every document, which took even more time.

Because they paid attention to every detail, it was difficult for them to see the “big picture” of the plot, and therefore this strategy turned out to be least effective of the different strategies, as mentioned by one participant:

P8: If I had to do it again, I'll scan through all documents several times until I get the big picture. This time, I read the documents much too carefully one by one and it took so long. I still haven't figured out what the story is about.

Example: P8 (Desktop setting)

Participant 8 opened each text file in order and read it carefully. He took notes on important events and names and identified them via the document number. After opening several documents in this manner, he decided to resize the documents and tile them on the right and left monitors so that he could see all of them. As he read more documents, he started grouping them by relevance on the monitors. He put documents related to certain locations on the left and the right monitor and any documents related to names on the top monitor. However, as he began to classify more documents and overlaps between locations and people became evident, he realized that this classification scheme was not working. Sometimes he highlighted important keywords within the text files. Throughout the analysis, he issued only eight search queries. Six of the eight queries were related to locations, such

as cities and states. After an hour, he finally started working on the task sheet on the main (middle) monitor and worked on the debriefing after another 20 minutes.

5.3 Strategy 3: Hit the Keyword (HTK)

Some participants used an unexpected strategy - an intensive keyword-based exploration. They did not begin the analysis by reading a specific document, but directly looked for a few specific keywords such as “terrorist” or “Al-Qaeda”. They read only the related documents and then searched for other terms that emerged during that time. This did not cover all of the documents, and these participants ignored the rest of documents that might not have been brought up.

Since the effectiveness of this strategy depended on the appropriateness of the terms chosen in the initial stage, performance varied across participants using this strategy. While P10 and P11 showed poor performance, P14 performed quite well using this strategy. He was in the Jigsaw setting, and he started using the List View even before he read any document or used the search control panel. He first added all documents in the first list, all people in the second list, and all places in the third list. Then he sorted the second list by frequency of appearance, which resulted in people appearing most frequently moving to the top. Selecting a person's name highlighted documents that contained the name in the first column, and he read those documents in the Document View. After reading a few documents relevant to those people who were at the top of the list, he moved to the third column, places, and repeated the same process. In this way, he was able to read most of the documents relevant to “important” people and places. This is a similar result to those who searched for particular names and places, but it was much more efficient in that he did not have to spend time in deciding which keywords to search and which documents to read. In fact, he made only four search queries total. In contrast, P10 and P11 made about 60 queries but only a few of them retrieved the most vital documents, which resulted in poor performance.

Example: P10 (Entity setting)

As the session started, participant 10 opened three Document View windows and placed them on the left, right, and the top monitors, and he placed the control panel (search box) on the center monitor. The participant entered the search term “Al-Queda” and this brought up a few documents. (In the post-session interview, the participant said that he found three Al-Queda members and traced their activities.) To read each document, P10 primarily used the Document View on the left monitor. He stated that he put Document View windows on the right and top monitors for future use. Next, he began to fill in the task sheet with names and events from the documents. He did not take separate notes. When the participant issued a new search query, he typically did not clear the previous documents, causing one Document View to have aggregated results of multiple queries. After one hour, he was still issuing queries for terms that were not relevant to the plot. In total, he read only 31 documents and missed many important ones.

5.4 Strategy 4: Find a Clue, Follow the Trail (FCFT)

The “Find a clue, follow the trail” strategy is a hybrid approach of the previous strategies, and four participants followed it. They invested some time in reading the first few documents to understand the context and find a clue, then followed the trail rigorously using search or other functionalities provided by the tool.

In theory, this may be a good strategy because the analyst’s attention is focused on relevant documents only. The initial investment in reading a few documents pays off because it increases the possibility of finding the right clue. The performance of participants who used this strategy is notably good.

When we more closely examined this strategy, we found two sub-strategies. While following the trail, P7 and P12 tried to read every document in the dataset at least once and made sure they did not miss any connections. This may work for a relatively small set of documents as was present here, but as the size of a dataset increases, an issue of scarcity of attention likely will arise because the analyst must keep track of what has been read and what has not.

Jigsaw participants P13 and P16, however, did not skim the rest of documents that were not in the trail. They read only 31 and 23 out of 50 documents, respectively. Since they gained the highest scores among participants, it seems clear that they focused only on important parts of the dataset, along the trail. From the log data we identified that both read all 23 important documents and that most of the documents irrelevant to the plot were not viewed. P16 identified one of the main players of the plot in the beginning of the analysis, and effectively explored the document collection following that person’s trail.

P16: I like this people-first approach. Once I identify key people, then things that are potentially important come up, too. I’m an impatient person and don’t want to read all documents chronologically.

This may be a fruitful strategy when there are a large number of documents. However, there still is a possibility of a dead-end if the analyst follows a wrong trail. In that case, the ability to quickly turn to another track is crucial.

P13: I started from a name and followed it. I had this one direction for a while and realized that it wasn’t a good way. I was kind of running into a dead end. Then I saw other names coming out over and over again, other areas coming out, then I got a story of what’s going on.

Example: P13 (Jigsaw setting)

Participant 13 started the analysis by skimming through a few documents in the Document View. In the beginning, he focused on a wrong story and his first six query terms were not relevant to the main plot. Once he realized that this path was not fruitful, he started the process again and turned his attention to other clues. In doing so, he noticed certain words such as “bombs” and other things relevant to arms and missiles. He pursued this angle, entered a search term in the control panel, moved to the Document View to read related documents, and then noticed other items coming up, expanding from there. Once P13 began to understand the story and key

information, he put separate Document View windows with related documents aside and moved them to the top monitor. He “froze” those documents by turning off event listening in these views. After about 30 minutes, he used the Graph View to validate his assumptions about relationships between names, events, and locations. P13 took notes throughout the session in conjunction with reading. His notes included not only key locations, events, and people, but also questions and hypotheses. About 50 minutes later, he started entering names on the task sheet. He said that he believed those names were important to the story and decided to put them together and re-read documents related to the names. He worked on the task sheet intensively while doing this iteration. At the end of the analysis, he revisited the Document View windows he had placed on the top monitor and referred to them while filling in the task sheet. He started writing the summary description with only 6 minutes remaining, but he already knew what he wanted to write and was able to finish it on time.

6 DISCUSSION

6.1 Jigsaw’s Influence

Among the four study conditions, the group using Jigsaw generally outperformed the other groups on the whole. The worst performance of a participant in this group was “good”, whereas the performance of participants in the other settings varied more. Based on observations, interviews, videos, and log analyses, we identified several benefits Jigsaw seemingly provided to users.

6.1.1 Supporting Different Strategies

Examining each participant’s analysis process, we note that the four Jigsaw setting individuals used three different strategies. This suggests that Jigsaw supported different analysis strategies well. For example, as discussed in the previous section, P14 was able to do keyword-based exploration effectively using the “sort by frequency” function of the List View. P15, who used the “overview, filter, and details” strategy, used the List View to grasp the main idea about important people, their organizations and locations after quickly going through all documents in the Document View. He opened an additional List View, put all the people and all the documents in two lists, and used it as an index when he needed to revisit documents in his second iteration. P13 and P16 both used the “find a clue, follow the trail” strategy, which was effective across settings. However, we found that these two individuals performed even better and more efficiently than those who used the same strategy in the other settings.

6.1.2 Showing Connections between Entities

Showing connections between entities such as people, organizations, and places is the key functionality of Jigsaw. We had a belief that showing connections would help the analysis process, and the study clearly revealed evidence to support this. Participants using Jigsaw performed well even though they did not fully take advantage of many system capabilities, partly due to limited training and unfamiliarity with Jigsaw. Mostly, they used the List View to explore

connections. Multiple participants in the non-Jigsaw settings wanted to see comprehensive connections between entities. Many of the generated notes contained important names, dates, and places. They were linked by lines and were used to assess the centrality of certain items, to understand what is important, and to decide what to examine further. The connections participants drew on paper and the functionalities they desired are similar to capabilities provided by Jigsaw. Figure 3 shows some examples. When asked about what were the most challenging aspects of the analysis, six out of twelve participants who did not use Jigsaw mentioned the difficulty in making connections:

P8: *Connecting names and documents was hard. Sometimes when two documents are related, there's no way to look it up if I hadn't marked [the connection] on each document.*

P3: *It was really hard to connect current information to what I read previously. Just too many names and places.*

Some participants also stated that they would change their strategy and make connections more visible if they had to do the task again:

P3: *I'd write down important names, places, and events and put them in different shapes by type and then show connections between them by lines and arrows.*

P5: *I'll use more of linking to make connections visible. It's not like mind maps or diagrams. I don't like maps without order. They don't help me understand the context. I'd prefer linking with some sort of ordering scheme.*

In contrast, none of the participants in the Jigsaw setting identified connections as an issue. Rather, they focused on the challenges in organizing and keeping track of relevant information.

6.1.3 Helping Users Find the Right Clue

Finding an appropriate clue early in the analysis is crucial and sometimes even determines the entire performance. Participants often seemed to take on a kind of “tunnel vision” about what was important, which may be problematic with large document collections. Even though the dataset used in this study was relatively small, participants still benefited from Jigsaw in finding a good starting point. Tag clouds, entity highlighting, and the connections in the List View helped to find the right clues:

P9: *Entity highlighting helped a lot. I didn't have to read all the documents and still could recognize the pattern.*

P15: *I think the tag cloud is really interesting. It was helpful to see some important terms when I did not know what it is about.*

P15: *I scanned all the documents first and got some rough ideas. But still I wasn't sure where to start. Then I opened the List View to explore connections between people and locations, and I started from there.*

6.1.4 Helping Users Focus on Essential Information

Even though analysts may find appropriate initial clues, it is still important to follow the trails in an efficient manner.

If relatively unimportant information diverts their attention, the investigative process may suffer no matter how quickly a good clue was discovered. We found that Jigsaw helped participants to follow the right trail and ignore irrelevant documents, thereby saving the participant's attention for important information. As we described earlier, two participants in the Jigsaw setting read only about half of the documents while the majority of other participants read all 50 documents at least once. These two Jigsaw setting participants (P13, P16) earned the two highest final scores. The other two participants (P7, P12) who used the same strategy and performed relatively well, in the Desktop and Entity settings respectively, both tried to read all the documents and keep track of other names or events while following the trail. This diverted their attention and hindered them from totally focusing on the main story. P12 (Entity setting) stated:

P12: *Because I searched for key phrases, read relevant stories, and went back to another document, I tended to lose track of all the dates that were going on.*

6.1.5 Reviewing Hypotheses

During analysis, the participants generated hypotheses about the hidden plot and gathered evidence that could support their hypotheses. Two of the Jigsaw setting participants found the Graph View to be useful as a confirmatory aid. P15 explored the dataset primarily using the Document View and the List View, and narrowed down to the most three important persons surrounding the plot. Then he used the Graph View to review his hypothesis and to check whether they were really key people in the plot, by quickly reviewing related documents and their connections to other entities.

P13: *Once I got some names and locations that I wrote down on paper, about one and half pages, I used the Graph View to get an idea of what's related. This is more confirmation rather than fact finding in that case. Everything in the middle was basically what I already knew about ... so ... I used it to validate what was going on. It was helpful but in a different sense. It's not about finding new facts but just asking like, was I right? What were things the graph is showing?*

6.1.6 Increasing Motivation

During the study, we noticed that participants in the Jigsaw setting often asked for more time, not because they had yet to discover the plot, but because they wanted to more closely examine their hypothesis and further evidence. Given that these individuals spent an additional 30 minutes in training prior to the session and the task time lasted for 90 minutes, we were surprised that three out of the four wanted to go on even longer. For the other settings, only P9 in the Entity setting asked for more time and her analysis was one of the worst. Three participants in the Paper setting stopped their analysis earlier than the allotted time. Their task performances were fair, very good, and excellent, so analytic performance did not seem to be a direct cause. These observations suggest that Jigsaw may have increased motivation and made the analytic processes more enjoyable.

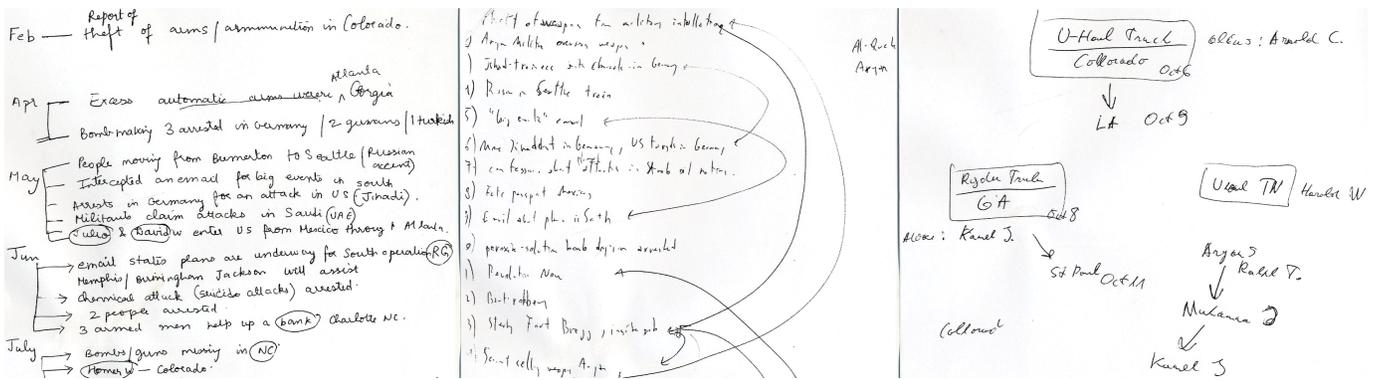


Fig. 3. Notes made by participants not using Jigsaw.

6.2 Observations on Sensemaking

Pirolli and Card have proposed a Think Loop Model of Sensemaking [14] consisting of two major loops, a foraging loop and a sensemaking loop, and several intermediate stages including “shoebox”, “evidence file”, and “schema”. According to the model, an analyst collects raw information into a “shoebox” through filtering and search. Snippets of this evidence are processed into an “evidence file”. Information from this evidence may be organized according to some “schema” or conceptual form either in the analyst’s head or with some external aid. This organization of information is used to marshal support for a story or set of hypotheses. Finally the information is transformed into a knowledge product, such as a briefing or a report. We observed study participants and how their actions related to this model.

6.2.1 Diversity in Sensemaking Processes

While the model is not linear and can proceed top-down or bottom-up with many loops, we found that the sequence of analysis significantly differed across individuals even in the same task with the same dataset. Some participants followed the sequence linearly with iterations; they extracted and jotted down important information while reading the documents, then organized the information according to a certain scheme such as time and location, eventually leading to a hypothesis. Some participants started organizing information as soon as they read documents, either by filling out the task sheet or drawing timelines/maps on paper, thus skipping the process of building an evidence file. Once they created a hypothesis, they took out snippets of information from the schema that did not support the hypothesis. Other participants immediately started from a hypothesis without the schema stage, and then worked on organizing to confirm the hypothesis. In this case, the schematizing stage took place at the end of the analysis process.

Individual differences also existed in each stage of the model. For example, the “read & extract” stage, in which evidence files are collected from the shoebox, exhibited individual differences. When encountering much unfamiliar information, it is not easy to extract nuggets of evidence simply by reading documents; the analyst usually needs some criteria to decide what to pull out. In our study, some participants started from a specific set of people and extracted information related to

those people. Those who used location as a criterion gathered all information related to specific cities or countries. Participants also extracted evidence files based on specific events such as arms thefts or truck rentals. Although participants used different approaches in this stage, it did not make a significant difference in the overall analysis process because the evidence files gathered tended to be similar regardless of the extraction criteria, as long as the analyst carried out the process thoroughly.

6.2.2 Power of Schematizing

It was the schematize stage that showed the most significant variance between individuals. During this stage, it seemed that each person had his/her own preferred organizational scheme such as a timeline, map, or diagram. For example, while most people wanted a timeline, the representations they envisioned were all different. Some people wanted a timeline organized by person and event; some wanted a timeline by location; others wanted a timeline categorized by story. Clustering was another organizational method employed by participants, but the classification scheme varied - by organization, by location, and by connectedness. The variances in this stage seemed to affect the entire analysis performance.

The time at which a participant first reached the schematize stage and how much effort the participant invested in this stage significantly affected the performance. When we further examined those who performed well independent of the setting, we found a commonality that all of these people spent considerable time and effort in organizing information. Most people used the task sheet as a tool for gathering their thoughts since the task sheet was structured by certain schemes (e.g., people, events, and locations). During the interviews, many participants explicitly described how completing the task sheet helped their sensemaking process.

P9: *Filling out the task sheet - all the events by date - was really helpful. At first, I started from people’s names, but at some point I jumped to the events instead of names, and found that much more helpful to make sense of the story. Jotting down didn’t help that much.*

P14: *Also the task sheet helped me a lot to make sense of all different stories. Being able to write down simultaneously while viewing the screen was very helpful. Actually, I found myself pondering a lot during that process.*

As the quotes indicate, participants did not expect the task sheets to help their investigation at first, but they noted the sheets' usefulness at the end. Note, however, that the participants were simply marking down entities from the documents on the task sheets, not new or meta information. The key difference was that the entities were being organized under a particular point of view.

Those participants who did not build schema or undertake some organizational process performed poorly on both task sheets and debriefing. Some of them did take a fair amount of notes, but no process of organizing the notes followed. Simply rewriting information without imposing an organizational scheme did not appear to help the sensemaking process.

6.2.3 Insight Acquisition

It is still difficult for us to identify exactly when people gained a key insight during the investigative process. When we asked the participants how they knew they were progressing towards the goals, the common answer was "when the pieces of a puzzle started being connected and put together." Rather than a spontaneous insight occurring (the "light bulb going on"), insight seemed to form continuously throughout the investigation, not unlike that described by Chang et al. [17]. Participants had difficulties identifying when they "got" the plot. P13 who gained the highest score, when asked about this difficulty, stated:

P13: Well, that's interesting. I don't know. Names coming up a lot, there's all these relationships like, for example, there seems to be Colorado and Georgia, and there were organizations there. You have this idea that just validates itself.

6.3 Comparison between Professional Analysts and Student Participants

Chin et al. [11] conducted an observational case study with professional intelligence analysts and revealed various characteristics of the analytical processes of the analysts. The goal of our research was similar – we aimed to understand how people would approach investigative analysis using visual analytic systems. While the ideal scenario would have been using real analysts as study participants, we conducted the study with graduate students for practical reasons. Also, we wanted to see to what extent our findings mirrored the analytical processes of real analysts. Comparing and contrasting findings from the two studies would be interesting in that most user studies in this area are performed with those who are not necessarily domain experts.

6.3.1 Common Characteristics

Chin et al. found that even for intelligence analysts (IAs), the analytical process is often seen as indefinite and self-perpetuating. Thus, IAs are always moving to new queries and new lines of investigation. Intelligence analysis is typically conducted within a specific period of time and to IAs, analysis is never fully complete but rather is valid as far as the current evidence and time of analysis shows. In our study, participants often expressed such uncertainty in their findings after the

session. Even those who earned the highest ratings explicitly said that they were still confused and not sure about their conclusions.

Due to a limited timeline, the IAs expressed that they would often abandon a systematic approach to satisfy time constraints. Under time pressure, IAs may mentally conduct many aspects of an analysis without adhering to a particular investigative path or documenting intermediate findings. Our student participants showed similar patterns. Even when they initially took a systematic approach, they soon gave up because of time constraints and instead mentally performed the investigation until they started working on the deliverables.

Two interesting characteristics of our study were how participants in non-Jigsaw settings tried to group documents into clusters by physical grouping and how they tried to make connections by drawing lines and relating relevant information on paper. The IAs took similar steps in their analytical processes. Chin et al. observed "one analyst sat on the office floor and laid out the hardcopies in a circle around her such that she could see the documents all at once. She labeled each of the documents to track them and piled them according to their types. She then manually drew a graph that showed the relationships among documents. Later, the IA redrew the document relationship graph in MS PowerPoint." Another IA read through all the documents and then marked each one with a number identifying its relevance and importance. He then physically ordered the documents in a pile from most to least important and relevant. He called this process "triage," and this "triage" appeared in many participants' analytical processes in our study as well. Once these documents were ordered, the analyst would extract key facts from one document at a time and insert them into a spreadsheet.

After IAs gathered, collected, and filtered evidence from materials, they examined the data to identify relevant patterns and trends. They primarily sought times, dates, and activities, and they searched for key relationships or connections among facts and evidence. We found a similar behavior in our student participants. It seemed that both analysts and student participants put effort into identifying important connections surrounding times and activities, although our participants also focused heavily on people.

Finally, Chin et al. found that an important step in the analysis process is to eliminate data and information that are deemed irrelevant in order to narrow the scope of the investigation and concentrate on the more essential details and facts. Eliminating irrelevant data was also one of the main findings from our study and we highlight how important it is to make the analytical process more efficient.

6.3.2 Professional Analysts' Unique Characteristics

Although we found that the studies with real analysts and with students identified similar findings and implications, it was evident that we could not capture every aspect of analytical processes of intelligence analysts within our study.

While our student participants did not have any specific analytical strategy prior to the analysis, IAs are knowledgeable about several strategies and use one they prefer. In the Chin et al. study, one common intelligence analysis strategy

used by the IAs is the analysis of competing hypotheses [18]. Using this strategy, an analyst delineates all possible hypotheses and then maps each piece of data to each hypothesis to assess whether the data supports, counter-indicates, or is irrelevant to the hypothesis. Another IA in the Chin study argued that intelligence analysis should not begin with any preconceived hypotheses. This analyst suggested letting “the data prove itself and suggest itself.” Yet another IA promoted a strategy of investigating the four specific areas of access, intent, motivation, and capability.

Interestingly, for IAs, physical files and folders still play prominent roles in the collection and organization of information. Despite receiving all case information in electronic form, each of the IAs printed out hardcopies of the individual documents. In their follow-up interviews, Chin et al. discussed that all of their analysts at one point had conducted intelligence analysis solely using physical files and folders. The use of traditional media continues to be a part of their analysis processes. In our study, however, participants performed the analysis almost solely with computational tools except for those in the Paper setting. We assume that the relatively short time (90 minutes) of the study imposed pressure and they might have felt that using computer support rather than traditional media would shorten the task time.

One notable difference between the two studies related to credibility and ambiguity of data. In the context of intelligence analysis, statements are not necessarily indisputable truths. IAs must confront information that may be ambiguous, incomplete, and/or conflicting. The level of credibility and validity is often judged by the credibility of the information source, the circumstances under which the data was collected, and corroboration by other sources. Data sources are also judged to have different levels of credibility and the corroboration of information is an ongoing activity during analysis. However, our student participants who neither received any training nor had any domain experience were not suspicious of the credibility of data itself and the data sources. Instead, they assumed that all information they received was accurate and factual.

Finally, IAs heavily rely on and exercise their personal knowledge when defining and establishing relationships, whereas the student participants had relatively little background knowledge to apply. In addition, our study did not address other important characteristics of IA processes such as how they collaborate, the form of their analysis products, and conventions of their everyday work practices.

6.4 Study Limitations

The study had several limitations that likely affected our findings. First, it was clear to us from the results (scores) and observations that there was quite a bit of variability among the participants. We speculate that certain individuals simply have better innate skill at such analysis tasks. Hopefully, visual analytics tools can enhance people’s abilities whether they are weak or strong. In addition, our participants were graduate students, not professional analysts. As discussed in Section 6.3, student participants and professional analysts showed both commonalities and unique characteristics.

Though it would have been interesting to see if correlation between an investigative strategy and a setting exists, the small sample size (16) did not allow us to examine the relationship. For example, we could take the “Find a clue, follow the trail” strategy and see if a particular setting better supported that strategy compared to other settings. Examining that correlation yields 16 (4 settings x 4 strategies) experimental conditions and thus requires many more participants.

We compared Jigsaw to other traditional tools, but not to other visual analytics systems. Comparing the usage of our tool to other existing systems developed for investigate analysis would have generated more insightful findings and implications.

For the study, we used a relatively small document collection, which likely would not be the case in reality. The collection size was chosen to make the experiment feasible in a reasonable amount of time. We speculate that some of the findings would only be amplified when working with larger document collections. Throughout the discussion we identified numerous situations where larger datasets would place even more importance on highlighting connections, following evidence trails, and organizing data and evidence.

The analytic scenario used in the study was a *targeting scenario*, one in which analysts seek to “put the pieces together” and identify a hidden plot. Many investigative scenarios have no clear, specific solution, however. Instead, they involve general knowledge acquisition over a long time period. Developing evaluation strategies and measures for these scenarios appears to be particularly challenging.

It was clear to us that even with the two-phased training for Jigsaw, participants in that condition still overlooked many useful capabilities of the system. With further experience and training, we would hope that the system would be even more beneficial. In particular, the experience of performing an investigation like this appeared to place participants in a position where they could better understand system capabilities if given further training.

7 STUDY IMPLICATIONS

7.1 Design Implications for IA Tools

The study and its results suggest several design implications for visual analytics systems for investigative analysis.

Facilitate Clue-Finding

Study participants who employed the “find a clue, follow the trail” analysis strategy generally performed well overall. Thus, investigative analysis tools that support analysts in finding appropriate starting points or clues and then following the trail of these clues efficiently could be beneficial. Further, the performance of those participants who were able to focus only on relevant documents was outstanding. Investigative analysis tools should help direct the analyst’s attention to the most critical information.

Foster Smooth Transitions Between SM Stages

The study demonstrated that people do frequently move between stages of the Think Loop Model, particularly in the middle parts of the model. Investigative analysis tools should

allow smooth transitions between the “shoebox”, “evidence file”, and “schema” stages so that different sequences of the sensemaking process can be supported. Currently, the focus of Jigsaw is on the “shoebox” and the “evidence file” stages, but it lacks powerful support for the “schematize” stage. While Jigsaw does appear to help analysts finding nuggets of information effectively, it does not really support putting those pieces of evidence together. In other words, analysts may easily discover the pieces to be put in a puzzle and have a sense of which piece goes where, but they should also receive help in putting the pieces together. The ability to work on extracting evidence files and organizing them into a schema will significantly help the sensemaking process.

Support Evidence Marshalling

For Jigsaw to be a comprehensive investigative analysis tool, it is crucial for the system to include a workspace in which the analyst can simply drop/paste entities, draw connections between them, and add annotations, capabilities found in systems such as Analyst’s Notebook [19], the Sandbox [7], and Entity Workspace [5]. Several participants pointed out this missing functionality as well, including:

P16: *Remembering what I had already found was hard. Keeping track of names was really hard, too. When I was reading a document about truck rentals in different cities, I remembered I read a similar document before. Oh yeah, there was somebody who rented a truck from Chicago to Minneapolis but then I forgot his name and it was really frustrating.*

P12: *I’d probably do something like this [the task sheets] but either spread them out or do it on notepad to give me more room so that I can just cut and paste things and move things around.*

Allow Flexibility in Organizing

When supporting the “schematize” stage, developers of investigative analysis tools should consider that individuals will choose different organizational metaphors or schemes. For example, even for a timeline, individuals imagined many different types of timelines and they were quite insistent about this approach. Rather than providing one fixed schema, allowing flexibility and room for customization will be beneficial. One participant wanted to have the ability to organize a timeline by “story”, which also requires flexibility in organizational schemes.

P7: *It would be good to have categorized keywords or events with relevant people/activities sorted by time. For example, I can have multiple stories such as passport, truck rental, Al-Qaeda, things like that, and under each keyword, all related people/activities are listed in a sequential order.*

Tool developers may consider having a system suggest a few organizational schemes when the analyst has created a significant evidence file but still does not have a schema, particularly for novice analysts. Staying too long at the “evidence file” stage appears to impede the analysis process so suggestions of organizational schemes may be beneficial.

Suggest Alternative Paths but Support Task Resumption

It is not uncommon for an analyst to confront a dead-end or find evidence that refutes an existing hypothesis. Investigative analysis tools need to support the analyst to find appropriate next steps or alternatives by making the milestones of the investigative process explicit. In this way, the analyst can come back to the point where she/he was earlier and start over from that point. This also ensures that the analyst can proceed further without being too concerned about keeping track of past states.

P16: *I was managing too much information. While in the analysis, I was really afraid of [getting] out of track, so I didn’t want to go further at some point. I always kept coming back to the previous stage because I wanted to keep the main story line.*

7.2 Evaluation Implications for IA Tools

The study also suggested a number of ways to help evaluate investigative analysis systems. By comparing system usage to more traditional methods but otherwise giving participants freedom to perform as they wished, we feel that the findings are both realistic and provide ample grounds for contextual analysis and comparison.

We also suggest that the evaluation of investigative analysis tools focus on collecting more qualitative data. While quantitative data is useful when a solution is well-defined and measurable, the nature of investigative analysis is exploratory and flexible. It may be too limiting to assess the value of a system solely based on statistical results. Identifying best practices supported, particular pain points, and future design requirements can be better achieved through interviews and observations. When possible, we suggest using quantitative data such as usage log files and analysis scores to help understand qualitative results.

Findings from the study suggest potential questions to be answered in the evaluation of investigative analysis tools:

- Does the tool help to provide information scent appropriately, thus helping to find initial clues?
- Does it guide the analyst to follow the right trail, without distraction?
- Does it support different strategies (sequences) for the sensemaking process? That is, does it support smooth transitions between different stages of the model?
- Does it allow flexibility in organization?
- Does it help to find appropriate next steps when encountering a dead-end?
- Does it facilitate further exploration?

In this study, we identified and used several metrics, which are broadly applicable to evaluation of investigative analysis tools:

- The number of important documents viewed, relative to the entire collection
- When the analyst first started creating representations such as notes and drawings
- The quantity of representations created

We also suggest two possible metrics for evaluating investigative analysis tools:

- Amount of time and effort in organizing
- Amount of time the analyst spent in reading/processing essential information

8 CONCLUSION

While many researchers in the visual analytics community firmly believe that new visual analytics technologies can benefit analysts, showing that is the case is still a challenging proposition. Clearly, one necessary step is to compare the use of new technologies to existing, more traditional methods. We conducted an experiment comparing students performing an investigative analysis exercise under one of four conditions. While lacking the size and depth to identify statistically significant differences, the study nonetheless suggested that visual analytics systems such as Jigsaw can benefit investigative analysis. Two aspects of Jigsaw turned out to be particularly helpful: showing connections between entities and narrowing down the focus.

Beyond that, this research makes several contributions to the visual analytics community:

- It provides an experimental design and methodology that others can emulate and apply.
- It describes how participants used visualization for analytic benefit and how its absence amplified challenges and difficulties.
- It provides a description of four analytic strategies employed by participants in seeking to identify a hidden plot.
- It identifies a number of design suggestions and capabilities to make visual analytics investigative analysis systems more effective.
- It suggests new evaluation metrics and qualitative factors for conducting experiments on these types of systems.

Evaluation of visual analytics systems must progress in step with new technical development for continued progress. Understanding how and why systems aid analysts will help to inform future designs and research. Our study provides initial evidence and insight in this area, and sheds light on many challenging open questions.

ACKNOWLEDGMENTS

This research is based upon work supported in part by the National Science Foundation via Award IIS-0915788 and by the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001.

REFERENCES

- [1] C. Plaisant, J.-D. Fekete, and G. Grinstein, "Promoting insight-based evaluation of visualizations: From contest to benchmark repository," *IEEE TVCG*, vol. 14, no. 1, pp. 120–134, 2008.
- [2] J. J. Thomas and K. A. Cook, *Illuminating the Path*. IEEE Computer Society, 2005.
- [3] C. Plaisant, "The challenge of information visualization evaluation," in *AVI*, May 2004, pp. 109–116.
- [4] P. Saraiya, C. North, and K. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE TVCG*, vol. 11, no. 4, pp. 443–456, 2005.
- [5] E. Bier, S. Card, and J. Bodnar, "Entity-based collaboration tools for intelligence analysis," in *IEEE VAST*, Oct. 2008, pp. 99–106.
- [6] D. H. Jeong, W. Dou, H. Lipford, F. Stukes, R. Chang, and W. Ribarsky, "Evaluating the relationship between user interaction and financial visual analysis," in *IEEE VAST*, Oct. 2008, pp. 83–90.
- [7] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort, "The sandbox for analysis: Concepts and methods," in *ACM CHI*, April 2006, pp. 801–810.
- [8] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: supporting investigative analysis through interactive visualization," *Information Visualization*, vol. 7, no. 2, pp. 118–132, 2008.
- [9] J. Scholtz, "Beyond usability: Evaluation aspects of visual analytic environments," *IEEE VAST*, pp. 145–150, Oct. 2006.
- [10] Y. Kang, C. Görg, and J. Stasko, "The evaluation of visual analytics systems for investigative analysis: Deriving design principles from a case study," *IEEE VAST*, pp. 139–146, Oct. 2009.
- [11] G. Chin, O. A. Kuchar, and K. E. Wolf, "Exploring the analytical processes of intelligence analysts," in *ACM CHI*, April 2009, pp. 11–20.
- [12] A. Perer and B. Shneiderman, "Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis," in *ACM CHI*, April 2008, pp. 265–274.
- [13] B. Shneiderman and C. Plaisant, "Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies," in *BELIV*, May 2006, pp. 1–7.
- [14] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *International Conference on Intelligence Analysis*, May 2005.
- [15] A. Robinson, "Collaborative synthesis of visual analytic results," in *IEEE VAST*, Oct. 2008, pp. 67–74.
- [16] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *IEEE Visual Languages*, Sept. 1996, pp. 336–343.
- [17] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky, "Defining insight for visual analytics," *IEEE CGA.*, vol. 29, no. 2, pp. 14–17, 2009.
- [18] R. Heuer, *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, 1999.
- [19] i2 - Analyst's Notebook, <http://www.i2inc.com/>.



Youn-ah Kang is a PhD student in the School of Interactive Computing at the Georgia Institute of Technology. She earned her master's degree from the School of Information at the University of Michigan-Ann Arbor, specializing in Human-Computer Interaction. With emphasis on information visualization and visual analytics, her current research interests include document sensemaking and evaluation of interactive systems.



Carsten Görg is a postdoctoral researcher in the College of Computing at the Georgia Institute of Technology. He received a PhD in computer science from Saarland University, Germany in 2005. His research interests include visual analytics and information visualization with a focus on designing, developing, and evaluating domain specific visual analytics tools to support users in various analytical tasks. Görg is a member of the IEEE Computer Society.



John Stasko is a Professor and Associate Chair of the School of Interactive Computing at the Georgia Institute of Technology. He received a PhD in computer science from Brown University in 1989. His research interests are in human-computer interaction with a specific focus on information visualization and visual analytics. Stasko is a member of the IEEE and is on the Steering Committee for the IEEE Information Visualization Conference.