

Effects of Visualization and Note-Taking on Sensemaking and Analysis

Nitesh Goyal
Cornell University
Ithaca, NY, USA
ngoyal@cs.cornell.edu

Gilly Leshed
Cornell University
Ithaca, NY, USA
gl87@cornell.edu

Susan R. Fussell
Cornell University
Ithaca, NY, USA
sfussell@cornell.edu

ABSTRACT

Many sophisticated tools have been developed to help analysts detect patterns in large datasets, but the value of these tools' individual features is rarely tested. In an experiment in which participants played detectives solving homicides, we tested the utility of a visualization of data links and a notepad for collecting and organizing annotations. The visualization significantly improved participants' ability to solve the crime whereas the notepad did not. Having both features available provided no benefit over having just the visualization. The results inform strategies for evaluating intelligence analysis tools.

Author Keywords

Analysis, sensemaking, visualization, interface design

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

In crime analysis, detectives and other police personnel examine witness and suspect interviews, crime scene reports, coroner's findings and many other documents in order to detect an underlying pattern and identify a culprit [9]. Solving a crime requires the analyst to "connect the dots" by identifying patterns and links between facts across documents, time and space.

This process of analysis is one of *sensemaking* [13], in which analysts iteratively forage for relevant information, integrate that information into schemas or hypotheses that explain what they have found, and use these schemas to guide decisions. For example, a homicide detective must identify which documents are most relevant to solving the crime, pour through them to uncover key people, places, weapons, and motives, and uncover relationships among these entities. Finally, during the decision making phase, analysts choose hypotheses to act on. For example, a homicide detective might decide that the evidence points to a single culprit and recommend that he be arrested.

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

CHI 2013, April 27–May 2, 2013, Paris, France.

Copyright © 2013 ACM 978-1-4503-1899-0/13/04...\$15.00.

A variety of tools have been developed to improve the sensemaking process. For example, tools help analysts visualize and manipulate data at different levels of granularity to detect links between objects in large datasets and construct alternative hypotheses [7,8,14,15], as well as collect and arrange data and notes for later reference [1,12,16]. Typically, these tools have been evaluated informally, with a handful of users [e.g., 4], or with all features available simultaneously [e.g., 11]. While such studies are valuable, they do not shed light on the unique benefit or cost of any one single design feature to the analytical process, nor do they in most cases provide sufficient data for statistical testing of the effects of a given feature.

GOALS

The current paper aims to evaluate two features commonly available in analysis tools: a visualization of relationships among documents and entities, and a notepad that allows recording and summarizing information. Based on previous research emphasizing the benefits of these elements for the analysis process, we propose:

H1: Analysts with a visualization feature will be better able to solve an analysis task than analysts without it.

H2: Analysts with a note-taking feature will be better able to solve an analysis task than analysts without it.

It is less clear if these two features will work synergistically, leading to better performance than with either one alone, or clutter the interface and take away from focusing on finding and reading documents. We therefore ask:

RQ: What will be the benefits of providing both the visualization and note-taking features?

METHOD

To evaluate the value of the visualization and note-taking features, we asked participants to solve a crime problem in which evidence for a serial killer was hidden among various documents. We separately manipulated the presence or absence of the visualization and the note-taking features in an analysis tool prototype we created for this experiment (Fig. 1), resulting in a 2 by 2 between-subject design.

Materials

We adapted experiment materials from [2]. The crime case documents included two "active homicide cases" and six "cold cases" (unresolved past cases). The active cases included a cover sheet, four witness and suspect interviews

and a Coroner’s report. The cold cases included a summary of the victim, time, method, and witness interviews. Four of the six cold cases were “serial killer” cases, and they demonstrated similar crime patterns (e.g., killed by a blunt instrument). The key clue to naming the serial killer was hidden in one of the active homicide documents. Additional documents included city maps, bus route diagrams, crime statistics, and a police department organization chart. In summary, clues for the serial killer were hidden across 30 carefully constructed documents with 56 possible suspects.

Additional materials included training materials, a practice task, a MO (*modus operandi*) worksheet in which participants could indicate key details pertaining to the crime, and an empty suspect list. A post-task report sheet included spaces to indicate, for each crime, the prime suspect (if identified), any known attributes of the suspect, the victim(s), the MO, and the location. A post-task survey included demographic questions, task performance measures, 5-point scales for rating the tools and a comment section.

Analysis Tool

We designed a research prototype consisting of five primary panes that can be turned on/off independently (Figure 1). The *Visualization* feature (a) shows all the documents in the dataset that contain entities in common with the active document as edges between the document nodes. The thickness of an edge is based on the number of unique entities in common between the joint documents, using TF-IDF. Documents are color-coded based on the cases or categories to which they belong. In many analysis tasks seeing links between documents from separate categories is critical. For

example, noticing that two separate crimes in different times and geographical areas share similar patterns in the crime scene findings may raise a flag to search for a common culprit. This visualization was implemented using the open source Radial-Tree component of ProtoVis [6].

The *Notepad* (b) is a text editor where evidence can be imported from an open document as well as freely typed in. Highlighting important text found while reading a document automatically copies it into the notepad for easier access in the future. The notepad can also be used to jot down comments and hypotheses, and resized like other features.

The modularity of the system components allows testing the value of each component individually and in combination. White space replaced the area where one or two components were absent according to the experimental conditions. The tool was implemented using Java Applet and XML.

Participants and Procedure

Forty U.S. students (48% female) were recruited through campus flyers and paid \$22.50 for a 90-minute study. They were randomly assigned to one of the four conditions: Vis only, Notepad only, both, or none, resulting in 10 participants per condition. They first performed the practice task. Then, they were seated at a 25” monitor and trained on using the analysis tool with the features available in condition. Participants then proceeded to work on the homicide cases. They were told that they had one hour to solve as many cases as they could, but were not informed of the presence of the serial killer (an adapted hidden-profile task). Afterwards they completed the post-task report and survey.

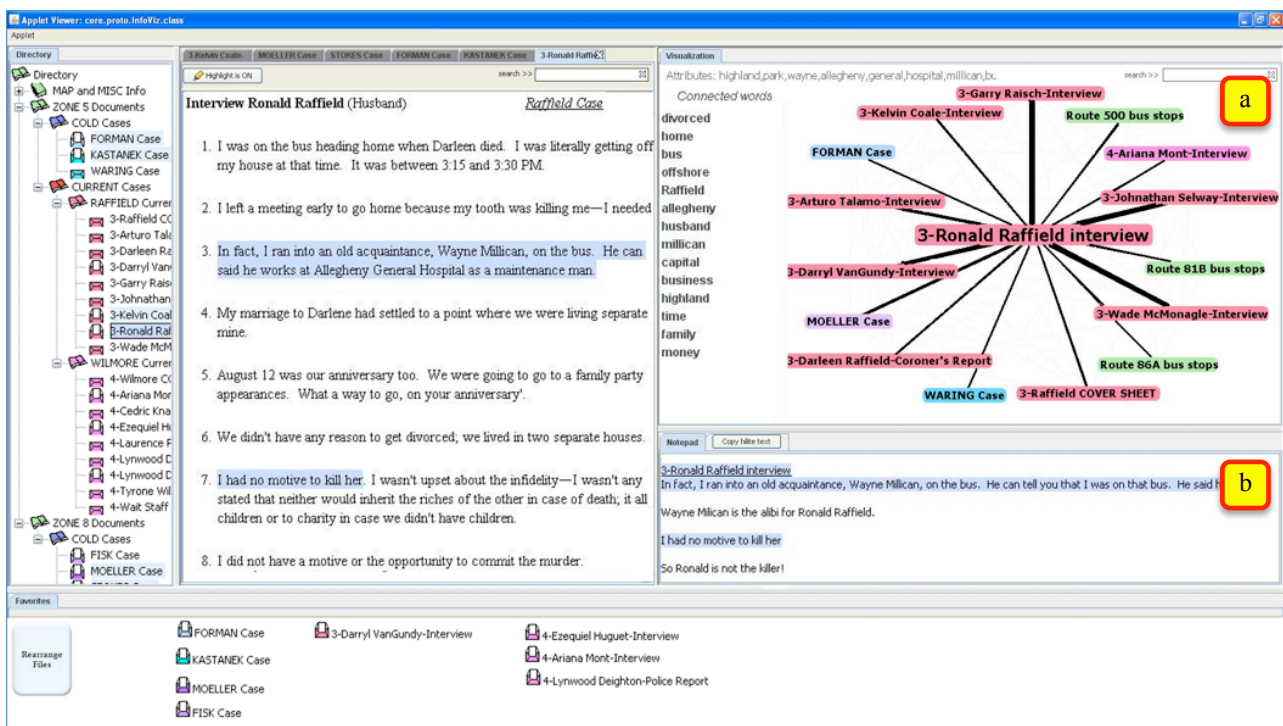


Figure 1. The Analysis tool prototype, showing the (a) visualization and (b) notepad manipulated in the experiment.

MEASURES

Performance. Based on the post-task report, participants received a score of 1 if they correctly identified the serial killer and 0 otherwise. We also counted how many key clues, out of 9, participants described in the post-task report. In the post-task survey, one question asked participants if they identified connections between the cold cases, and another question asked whether participants saw a connection between the cold cases and the active case holding the key clue. They received a score of 1 if they responded “yes” to both questions, a score of .5 if they responded “yes” to one question, and a score of 0 otherwise.

Feature use. Participants rated the usefulness of the interface features on five-point scales and added open-ended comments. We also approximated the time spent using each interface feature based on cursor location as indicated in the log files. We removed two values that were more than 2.5 SD above the mean, and divided by the total amount of time spent on the task to create percentage time measures.

RESULTS

Task Performance

Identifying the serial killer. Participants with the visualization (80%) were more likely to identify the serial killer than participants without the visualization (40%) (see Table 1, second column). In contrast, participants with the notepad were somewhat less likely to identify the serial killer (50%) than participants without the notepad (70%), and participants who had both the notepad and the visualization were less likely to identify the serial killer (70%) than those with the visualization alone (90%). In other words, it seems that the visualization improved task performance, whereas the notepad undermined performance. A binary logistic regression model, in which success was a dependent measure and visualization, notepad, and their interaction used as predictors, showed a borderline significant effect of the visualization ($B=2.20$, $S.E.=1.23$, $Wald=3.20$, $p=.07$), but no effect of notepad ($B=-.85$, $S.E.=.94$, $Wald=.82$, $p=.37$ and no interaction ($B=-.50$, $S.E.=1.57$, $Wald=.10$, $p=.75$). These results provide partial support for H1, and no support for H2.

Clue detection. Clue recall was highly correlated with successfully identifying the serial killer ($r[40]=.77$, $p<.001$). Participants detected 4.2 clues ($SD=1.8$) when they had the visualization, and 1.7 ($SD=1.8$) otherwise (Table 1, third column). They detected 2.7 clues when the notepad was available ($SD=2.3$), and 3.25 when not ($SD=2.0$). Clue re-

call scores were analyzed in a 2 (visualization) by 2 (notepad) ANOVA. The visualization led to significantly higher clue detection ($F[1,39]=19.43$, $p<.001$), the notepad had no effect on clues detection ($F[1,39]=1.12$, $p=.30$), and there was no interaction between variables. These results support H1, do not support H2, and show no specific benefit or detriment for using the visualization and the notepad together.

Relationship detection. The visualization significantly increased participants’ ability to detect relationships among documents, whereas the notepad had no effect (fourth column). A 2 (visualization) by 2 (notepad) ANOVA showed a significant effect of visualization ($F[1,39]=14.57$, $p=.001$), no effect of notepad ($F[1,39]=1.62$, $p>.20$) and no interaction ($F[1,39]=1.62$, $p>.20$). These findings, again, support H1, show no support for H2, and provide no evidence for benefit of both features together over one alone (RQ).

Use of Features

Time spent on features. When the visualization was available, participants spent about 11% of their time on it, and when the notepad was available they spent about 9% of their time on it (fifth column). Having both visualization and notepad did not influence the amount of time people spent on either one. Two one-way ANOVAs comparing time on the visualization with and without the notepad and time on the notepad with and without the visualization showed no significant differences (both $F[1,19]<1$, n.s.).

The amount of time spent on the visualization had no impact on how many clues participants detected. A one-way ANOVA using only participants who had the visualization showed that neither the presence vs absence of the notepad nor the percentage of time spent on the visualization affected the number of clues detected (all $F<1$, n.s.). Thus some other aspect of the visualization feature than the time spent on it helped participants find clues and solve the case.

Perceived feature usefulness. Participants found the tool features somewhat useful (see Table 1, rightmost column). When given both features, they found visualization more useful. Differences in ratings of the visualization and notepad were not statistically significant. The open-ended responses shed light on the benefits and downsides of these features. Some participants indicated that the visualization helped them understand how pieces of information were connected: “For just looking at one case, it isn’t very useful, but it is useful for trying to find links between cases” (P10, male). Others pointed out limitations of the visu-

Feature	Serial Killer Identification (%)	Clue Recall (0 to 9)	Relationship Detection (0 to 1)	Time Spent on Features (%)	Perceived Feature Usefulness (0 to 5)
Vis. Only	90 (SD=10)	4.5 (SD=0.34)	0.85 (SD=0.07)	12.61 (SD=1.97)	3.66 (SD=0.44)
Notepad Only	30 (SD=15)	1.4 (SD=0.47)	0.35 (SD=0.10)	8.64 (SD=2.82)	3.90 (SD=0.37)
Visualization and Notepad	70 (SD=15)	3.9 (SD=0.73)	0.60 (SD=0.10)	V:16.11 (SD=5.98) N:14.61 (SD=6.01)	V:4.30 (SD=0.21) N:3.70 (SD=0.42)
Control	50 (SD=16)	2.0 (SD=0.63)	0.35 (SD=0.10)	N/A	N/A

Table 1 Descriptive Summary (Means and Std. Deviations) across the four conditions for the five measures

alization in its current form: “it showed so many connections” (P20, female).

Similarly, some participants liked the ability to collect, organize and revisit pieces of information using the notepad: “to gather my notes and see and reread information that I had highlighted.” (P11, female). Others wanted more editing and sketching capabilities, as well as better contextualizing the notes in the documents: “scribble on or add comments to files” (P10, male). The notepad, as a simple text editor, might have lacked the richness of plain paper. These comments point to ways in which a note-taking feature could be better designed.

DISCUSSION

In this study we separately assessed the value of a visualization feature and a notepad feature in an analysis tool. As H1 predicted, the visualization helped people identify key clues for a serial killer and to point out the culprit. Counter to H2, the notepad provided no task performance benefits, alone or in combination with the visualization (in response to RQ), even though participants found it subjectively useful.

Observations of the experimental sessions suggest several explanations for *how* the visualization led to improved performance. Interacting with the visualization might have made it more obvious that there were connections among documents from separate crime cases, and thus potentially a single underlying culprit. In addition, the visualization may have given starting points for the investigation, shaping the paths participants took through the documents, or informed participants’ strategies for investigating the homicides.

Unlike full-fledged analysis systems, the notepad we implemented was not as sophisticated, which might explain why it did not improve performance. Another possibility is that while people successfully used the notepad to organize their thoughts, it slowed down their reading of the documents or reduced time spent developing schemas and hypotheses, leading to no net benefit for analysis.

One limitation of our study is that it used a simplified task with fewer documents and uncertainties than in most real-life crime investigations. Second, we focused on only two of the features that analysts would need. Third, the design of notepad was simplistic, and lacking a sketching capability, suggesting the low performance to be a function of this particular design, and perhaps not note-taking in general. Fourth, this experiment was a single session experiment to simulate time-sensitive sensemaking. However, in other situations, sensemaking might be a multi-session task, requiring extensive note-taking for archival [3]. We also did not test the value of tools for collaborative analysis [5,10].

CONCLUSION

This paper makes two contributions. First, we demonstrate the utility of testing individual features of complex analysis tools. Our experiment raises questions about the potential constraints on usefulness of different features of an analysis tool. Second, using a controlled experiment, we demon-

strate how specific features of the design facilitate or hinder the analysis task, how users perceive them, and how we can learn from mistakes to provide analysts with better tools.

REFERENCES

1. Andrews, C., Endert, A. & North, C. 2010. Space to think: large high-resolution displays for sensemaking. *Proc. CHI '10*, 55-64.
2. Balakrishnan, A.D., Fussell, S.R., & Kiesler, S. 2008. Do visualizations improve synchronous remote collaboration? *Proc. CHI '08*, 1227-1236.
3. Boch, F, Piolat, A. 2005. *Note taking and learning: a summary of research. WAC Journal '05*, 101-113.
4. Bier, E.A., Card, S.K., & Bodnar, J.W. 2010. Principles and tools for collaborative entity-based intelligence analysis. *IEEE TVCG*, 16(2), 178-191.
5. Borge, M., Ganoë, C.H., Shih, S-I & Carroll, J.M. 2012. Patterns of team processes and breakdowns in information analysis tasks. *Proc. CSCW '12*, 1105-1114.
6. Bostock, M., & Heer, J. 2009. ProtoVis: A graphical toolkit for visualization. *IEEE TVCG*, 15, 1121-1128.
7. Chin, G. Jr., Kuchar, O.A., & Wolf, K.E. 2009. Exploring the analytical processes of intelligence analysts. *Proc. CHI '09*, 11-20.
8. Convertino, G., Billman, D., Pirolli, P., Massar, J.P., & Shrager, J. 2008. The CACHE study: Group effects in computer-supported collaborative analysis. *Proc. CSCW '08*, 353-393.
9. Gottlieb, S., Arenberg, S., & Singh, R. 1994. *Crime analysis, from first report to final analysis*. Santa Barbara, CA: Alpha Publishing.
10. Heer, J. & Agrawala, M. 2008. Design considerations for collaborative visual analytics. *Info. Vis.*, 7, 49-62.
11. Kang, Y-a, Gorg, C., & Stasko, J. 2011. How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE TVCG*, 17, 570-583.
12. Pioch, N.J. & Everett, J.O. 2006. POLESTAR: Collaborative knowledge management and sensemaking tools for intelligence analysts. *Proc. CIKM '06*, 513-521.
13. Pirolli, P. & Card, S. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proc. International Conference on Intelligence Analysis*, 2-4.
14. Shrinivasan, Y.B. & van Wijk, J.J. 2008. Supporting the analytical reasoning process in information visualization. *Proc. CHI '08*, 1237-1246.
15. Stasko, J., Goerg, C., & Liu, Z. 2008. Jigsaw: Supporting investigative analysis through interactive visualization. *Info. Vis.*, 7, 118- 132.
16. Wright, W., & et al. 2006. The sandbox for analysis: Concepts and methods. *Proc. CHI '06*, 801-810.