Do Collaborators' Annotations Help or Hurt Asynchronous Analysis?

Ruogu Kang
HCI Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA

ruoguk@cs.cmu.edu

Sara Kiesler
HCI Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA

kiesler@cs.cmu.edu

Abstract

Our study investigated the use of annotations in an asynchronous crime-solving task. In Study 1, regardless of whether they anticipated a partner, participants had better performance if they annotated more about connections across documents. In Study 2, annotations that pointed to more connections across documents improved the performance of the second participant. Annotations that pointed to few connections across documents hurt performance, especially when people were more aware of their partners. This research suggests that future collaborative tools should help people discern useful from useless annotations.

Keywords

Collaborative analysis, asynchronous collaboration, annotation, problem solving, sensemaking

ACM Classification Keywords

H5.3. Group and Organizational Interfaces: Asynchronous interaction, Computer-supported cooperative work.

## General Terms

Experimentation, Human Factors

INTRODUCTION

The discovery of Osama bin Laden, of a serial killer (www.trutv.com/library/crime/serial_killers/predators/robert_pickton/1.html), and of the mysterious origin of e coli bacteria in spinach all relied on the work of many analysts working collaboratively on large amounts of information. Investigative analysis is often too complex and too distributed for the capabilities of individual analysts working alone (e.g.,[5]). As in the examples above, analysts often work asynchronously, and may even gain from doing so by combining their varied perspectives and doing more complete analysis than if they had to work synchronously [6]. Researchers have proposed new approaches to improve asynchronous collaborative analysis [3,8]. For instance, existing tools can aid in visualizing collaborators' actions [2], or enable commenting and bookmarking on shared visualizations
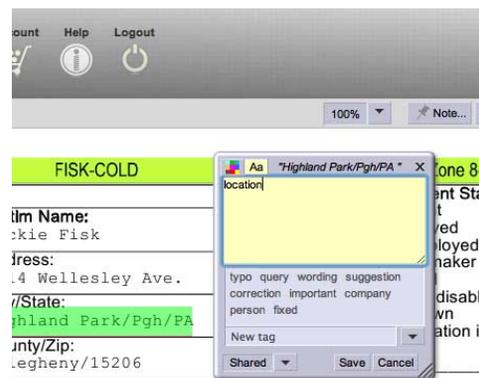
Figure 1. A screenshot of the annotation interface

[3]. We focus in this paper on the value of annotations that allow one collaborator to share his thinking and line of evidence with another collaborator. Shared annotations may be helpful because they point to others' attention and interpretations of data, and provide a record of activity on documents[7]. On the other hand, considerable research suggests that others' opinions and judgments can misdirect people[1,4]. If annotations call attention to irrelevant or misleading data, they may interfere with effective analysis.

The purpose of our research was to investigate the effects of annotations in more detail than has been achieved in prior work. We asked whether annotations would support or impede the work of analysts and their partners. To do so, we experimentally decomposed the asynchronous collaborative analysis process into two separate studies. We examined the quality and types of analysts' annotations and their effects on task performance in study 1, and the effects of a collaborator's prior annotations on the second analyst's performance in study 2. Through this work we show how and why annotations vary in their usefulness, and whether annotation is likely to help or hurt collaborative asynchronous investigative analysis.

Study 1: How analysts annotate documents

Study 1 had 48 participants, all undergraduate or graduate students at a private university. Each participant was paid $15 for participating in the experiment for 1.5 hours. Participants were given one of two crimes to solve. Participants on the serial killer task force were told their task was to identify a probable serial killer from 21 documents including seven homicides that occurred in the same city. Seven of the 21 documents included critical evidence, such as a

similar cause of death by blunt instrument. The other crime problem was about a serial robber, which has similar structure to the serial killer paradigm. All documents participants used were uploaded to a website (http://a.nnotate.com/). Participants were given brief instructions on how to annotate documents using the tool provided by the website (as shown in Figure 1). After one hour, participants were asked to complete a progress report in which they were to write down a preliminary suspect and evidence they had found to support this claim.

We randomly assigned participants to two experimental conditions. In the Partner condition, they were told to leave annotations for a partner who would work on the same assignment after their session ended. In the No Partner condition, participants were informed that they would be asked to come back to work on the same assignment again. They were told they should leave their annotations for their own future retrieval.

Results of Study 1
First of all, we inductively categorized annotations according to the topics participants mentioned.

Annotations created by participants were coded into four categories of topics as shown in Table 1 on next page. "Evidence" and "suspect" categories referred to annotations that explicitly mention any evidence or suspicious individuals' name. The "Connective" category included all annotations that pointed to linkages across two (or more) documents, such as the fact that crimes occurred in similar locations, or that there were contradictory statements from two reports. The "Pattern" category included annotations that remarked on a repeating pattern or an anomaly in a pattern.

Category	Example
1. Evidence	
Motive	"reduced salary… could be a motive"
Weapon	"handgun"
Opportunity	"cleaning staff has access to the building"
Time, location & others	"the crime happened around midnight"
2. Suspect	"Mr. Talamo is possible suspect"
3. Connective	"same as Fisk case"
4. Pattern	"same weapon as other robberies"

Table 1. Annotation topic categories
(inter-rater agreement kappa coefficients between two coders are .84, .89, .86, and .91 for evidence, suspect, connective, and pattern)

Among all 48 participants, 37.5% of them correctly named the serial suspect in their progress report, and 62.5% of them failed to identify the correct serial suspect. Using the annotations categories in Table 1, we performed logistic regression analysis to find out which type of annotations better predicts whether people can identify the correct suspect or not in the report. The predictor variables in the models are (a) having a partner (b) assignment to serial killer or robbery task force crime, (c) number of annotations in category of interest, and (d) number of annotations in other categories. The results showed that, controlling for partner condition and task assignment, annotations that pointed to connections across documents were the most significant predictor of better performance (X^2 [1,43] = 9.81, p < .01 for identifying the correct suspect). The more connective annotations people created, the better they performed in the task.

We then tested the effects of the Partner vs. No Partner conditions on creation of annotation and their task performance. There were no differences between conditions in the annotations people created, suggesting either that participants did not attend to the purported partner or else created similar annotations for themselves as for a partner. The partner conditions did not show a difference in task performance as well.

Study 2: Are those annotations helpful?
Study 2 had 60 participants, all undergraduate or graduate students at a private university. Participants were randomly assigned into three conditions in study 2: Many Connective Annotations (6 of 12 annotations on their documents pointed to connections between documents), Few Connective Annotations (2 of 12 annotations on their documents pointed to connections

between documents), and a No Annotations control group. The annotations in the first two conditions were created from samples of annotations created by participants in Study 1 who had a partner and worked on the serial killer assignment. The procedure for doing the task was similar to the procedure used in Study 1 except that participants in the first two conditions saw annotations from a prior partner. Participants in all three conditions were informed that their partner had worked on the same problem before their session. We added two new measures of task performance by conducting a survey after the task, which included questions relating to clues about the serial killer and questions about suspicions participants might have had. Also, we asked participants whether they were aware of their partner.

Results of Study 2
Similar to Study 1, we conducted a logistic regression analysis for identifying the correct suspect and linear regression analyses for the number of correct clues remembered and number of correct suspicions reported. Although we did not find an overall main effect of condition on the proportion of participants who identified the correct suspect, the number of correct suspicions reported by participants was found to be significantly different across conditions (F(2,55) = 4.67, p = .01). Figure 2a shows that those in the Many Connective Annotations condition outperformed those in the other two conditions (p < .05). The effect of conditions on remembering clues was in the same direction, but not significant. In addition, we found that awareness of the partner interacted with the impact of annotations on the number of correct clues remembered (interaction F (2,52) = 3.20, p = .05) as shown in Figure 2b. Here, participants who were more aware of their partner in the Few Connective Annotations condition were marginally
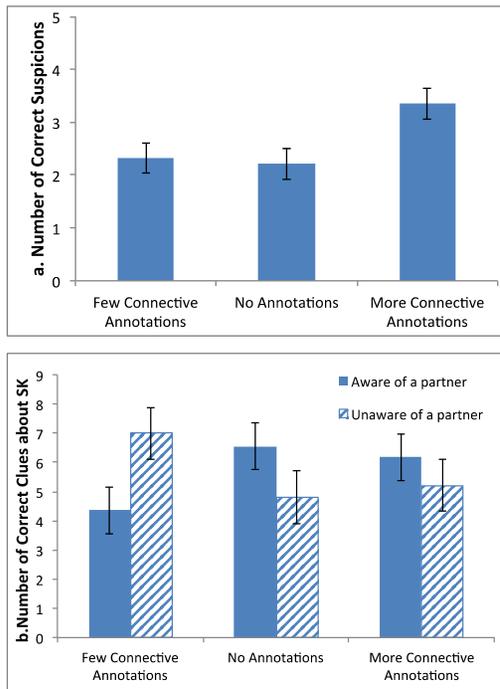
Figure 2. The effect of partner's annotations on the number of correct suspicions detected (Figure 3a), and the interaction effect of partner awareness and annotation conditions on the number of correct clues reported (Figure 3b)

worse at remembering the correct clues than those who were unaware of their partner in the same condition (t (17) = 2.06, p = .06). In the same figure, those in the No Annotations and Many Connective Annotations conditions who were more aware of their partner did better than those who were unaware. The same pattern held on the number of correct suspicions detected (interaction F (2, 52) = 3.00, p = .06). These results suggest that participants who paid attention to their failed partners were led in the wrong direction.

Conclusion and Future Work

We did not find any effect of expecting a partner on the creation of annotations in Study 1, but we did find strong differences in how participants annotated their documents and in how different annotations affected a subsequent partner's performance in the task. The results of Study 1 show that higher quality annotations pointed out connections across documents. Our results from study 2 verify that these annotations had a significant impact on participants' performance. Viewing higher quality annotations helped participants perform better in identifying the correct suspect and making sense of the problem than those who received poorer annotations. In addition, when participants were unaware of a partner whose annotations were poorer quality, these participants tended to perform better than those who were aware.

This study comprises a first step in experimental work on asynchronous collaborative analysis. Future research will need to address collaborative analysis tasks when partners work sequentially or reciprocally, examining how they cooperate to perform

asynchronously across time and geography. To answer our research question, whether collaborators' annotations help or hurt asynchronous analysis, we suggest that the helpfulness of annotation is conditionally based on the quality of annotations provided by analysts and also on their perception of these annotations. We propose using computational methods such as natural language processing to elicit higher quality annotations that connect evidences across documents, and using visualization techniques to direct collaborators' attention to higher quality annotations.

REFERENCES

[1]   Balakrishnan, A.D., Fussell, S.R., Kiesler, S., and Kittur, A. Pitfalls of information access with visualizations in remote collaborative analysis. In *Proc. CSCW 2010*, (2010), 411-420.

[2]   Gotz, D. and Zhou, M.X. Characterizing users' visual analytic activity for insight provenance. *Information Visualization* 8, 1 (2009), 42-55.

[3]   Heer, J., Viégas, F.B., and Wattenberg, M. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM* 52, 1 (2009), 87–97.

[4]   Hullman, J., Adar, E., and Shah, P. The Impact of Social Information on Visual Judgments. In *Proc. CHI 2011*, (2011), 1461--1470.

[5]   Kiesler, S. and Sproll, L. Group decision making and communication technology. *Organizational Behavior and Human Decision Processes* 52, 1 (1992), 96-123.

[6]   Lojeski, K.S., Reilly, R., and Dominick, P. The role of virtual distance in innovation and success. In *Proc. HICSS'06*, (2006), 25c-25c.

[7]   Millen, D.R., Feinberg, J., and Kerr, B. Dogear: Social bookmarking in the enterprise. In *Proc. CHI 2006*, (2006), 111-120.

[8]   Willett, W., Heer, J., Hellerstein, J., and Agrawala, M. CommentSpace: structured support for collaborative visual analysis. In *Proc. CHI 2011*, (2011), 3131-3140.