# CrowdTrace: Visualizing Provenance in Distributed Sensemaking

Tianyi Li*† Yasmine Belghith† Chris North† Kurt Luther†

*Loyola University Chicago †Virginia Tech

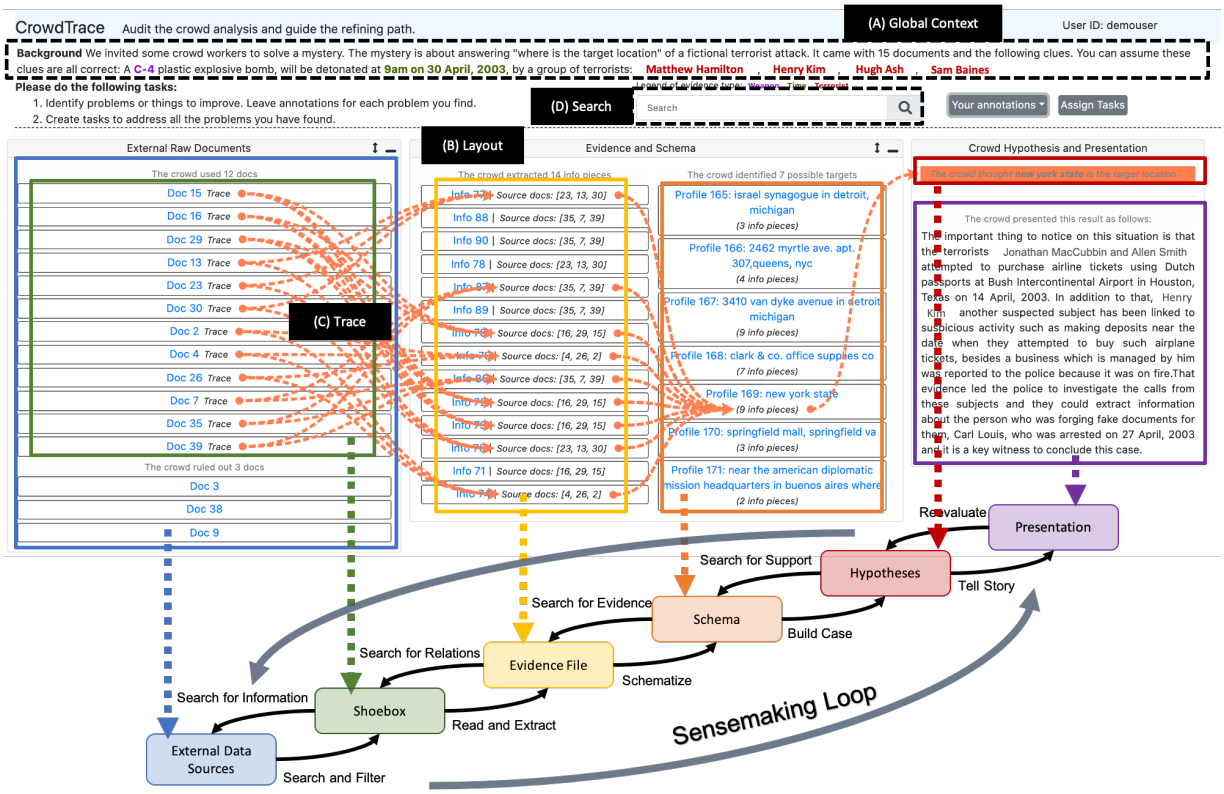


Figure 1: The auditing interface of the CrowdTrace system, annotated to show correspondences to the Sensemaking Loop theory. (A) Global context is always on the top of the interface. The (B) layout of the crowd analysis represents the analysis provenance structured by the sensemaking loop [22]. The Trace buttons allows users to (C) trace the information flow. Users can also (D) search for keywords in the crowd analysis; all the occurrences will be highlighted and the corresponding documents will be expanded.

## Abstract

Capturing analytic provenance is important for refining sensemaking analysis. However, understanding this provenance can be difficult. First, making sense of the reasoning in intermediate steps is time-consuming. Especially in distributed sensemaking, the provenance is less cohesive because each analyst only sees a small portion of the data without an understanding of the overall collaboration workflow. Second, analysis errors from one step can propagate to later steps. Furthermore, in exploratory sensemaking, it is difficult to define what an error is since there are no correct answers to reference. In this paper, we explore provenance analysis for distributed sensemaking in the context of crowdsourcing, where distributed analysis contributions are captured in microtasks. We propose *crowd auditing* as a way to help individual analysts visualize and trace provenance to debug distributed sensemaking. To evaluate this concept, we implemented a crowd auditing tool, *CrowdTrace*. Our user study-based evaluation demonstrates that CrowdTrace offers an effective mechanism to audit and refine multi-step crowd sensemaking.

**Keywords:** Crowdsourcing; sensemaking; crowd auditing

## 1 Introduction

During sensemaking processes, the data undergoes multiple transformations as it moves through different stages of human reasoning. Analyzing the provenance is important for verifying prior insights, scaffolding collaboration, and supporting communication in a range of domains and contexts [23]. A vast majority of visual analytics tools provide provenance support through capturing behavioral histories or manual annotations [30]. However, these tools can fall short in distributed sensemaking, which leverages prior insights without direct collaboration among analysts [7]. Futhermore, the complex sensemaking process of multiple iterative loops make asynchronous communication and hand-off difficult. Problems and errors in intermediate analysis get compounded and are hard to trace back [17]. Refining the results of distributed sensemaking requires analysts to understand the provenance and provide actionable feedback.

One ideal context to study distributed sensemaking is crowdsourc-

*e-mail: tli@cs.luc.edu
†e-mails: byasmine@vt.edu, north@cs.vt.edu, kluther@vt.edu

ing, where all of the analysis actions are captured as microtasks. Crowdsourced sensemaking has been applied to difficult sensemaking problems in various application domains including solving mysteries [16], online shopping [12], and Q&A sites [10]. The VIS community has also used crowdsourcing to evaluate visualizations. In this paper, we explore the inverse of that tradition and use visualization to diagnose the crowd sensemaking provenance.

We present a novel concept, *crowd auditing*, as a way to help individual analysts visualize and trace provenance to debug distributed sensemaking. Rather than improving crowd performance in a single step, the goal of crowd auditing is to probe the problems across multiple crowd processes, and steer the refinement with a top-down approach. To support crowd auditing, we developed *CrowdTrace*, a software prototype that visualizes the crowd analysis provenance and supports auditors in identifying problems and providing feedback to refine the crowd results. We evaluated CrowdTrace with a user study of 19 participants. Each participant was asked to audit a pipeline of crowd analysis for solving a mystery. The crowd analysis is formulated as a bottom-up path of the sensemaking loop [22], which represents a broad class of sensemaking tasks. Using CrowdTrace, participants successfully identified important problems in the crowd analysis and provided actionable feedback by creating high-quality microtasks for crowdworkers to address the problems.

In summary, this paper makes two main contributions: 1) a novel approach, *crowd auditing*, to address the challenge of mixed-quality results in multi-step crowd sensemaking; and 2) a crowd auditing system, CrowdTrace, that visualizes crowd sensemaking provenance and supports top-down refinement of crowd analyses.

## 2 Related Work

### 2.1 Analytic Provenance and Visualization Support

The steps that an analyst takes to make sense of raw data can shape the insights discovered, and are often considered as important as the final product [21]. Visual analytics tools capture interaction logs, intermediate visualizations, or user annotations to preserve the *analytic provenance* of sensemaking processes [23]. Visual analytic tools with provenance support have been built to help intelligence analysts to reason with different types of evidence in their sensemaking processes [30]. North et al. [20] proposed five interrelated stages of examining analytic provenance. First, understanding how the information was presented and *perceived* by the user [4]. Second, identifying the most appropriate representation of user reasoning processes (e.g. interaction logs [23] or user annotations [29]) *captured* by the visualization system. Third, *encoding* the captured provenance in a pre-defined form. Fourth, make sense of and *recovering* the provenance. Lastly, *reusing* the insights in new data or domains. In this work, we build on these five stages and explore provenance analysis in crowdsourced sensemaking, where the reasoning process is distributed and not captured a unified visual interface.

### 2.2 Crowdsourced Sensemaking

Crowdsourcing provides an ideal context to study distributed sensemaking as it can aggregate human intelligence at a large scale through micro contributions. Researchers usually need to decompose a complex problem into workflows to support crowdsourced analysis. For example, Crowd Synthesis [1] scaffolds expertise for novice crowds via a *classification-plus-context* approach, where crowds first re-represent the text data and then iteratively elicit categories. CrowdIA [16] enables novice crowds to solve mysteries with a pipeline adapted from the sensemaking loop [22]. However, the crowdsourced analysis cannot be perfect. The errors in one step of the crowd process can propagate to later analysis and affect the final outcomes [17]. Better task design can reduce crowd errors but is nontrivial and usually requires multiple rounds of refinement [15]. Furthermore, exploratory analysis requires impromptu adaptation of the sensemaking process, which further challenges pre-determined workflows [24]. Visualizing crowd outputs can help reveal redundancies [31] and low-quality work [26], but they are not able to trace the error propagation nor refine the crowd analysis. In this work, we explore how to visualize and trace the crowd sensemaking provenance to debug and refine the crowd's analysis.

### 2.3 Providing Feedback in Sensemaking

Feedback is an important discovery pathway in sensemaking processes to correct different kinds of errors [13]. The data-frame model of sensemaking [14] describes feedback as "discovering inadequacies of initial account, comparison of alternative accounts, reframing the initial account and replacing it with another." The sensemaking loop model [22] involves top-down processes that test theories against the data to validate the analysis outcome in each sub-process. The most common sources of feedback in sensemaking processes are clients [22] and peers [28]. Peer feedback generally involves analysts who share the same problem-solving goals inspecting one another's work [3]. Self-assessment has also proven useful, achieving comparable results to external sources of feedback [5]. In collaborative sensemaking, analysts need to understand what each person has done to effectively coordinate their efforts. Xu et al. enabled a single analyst to review and analyze previous chart-driven analyses using a meta-visualization approach [32]. Knowledge Transfer Graphs [33] automatically capture, encode, and streamline analysts' interactions to support hand-off of partial findings during analysis. *Dimension coverage* [27] helps analysts understand the analysis history of previous collaborators and facilitates continuous coordination of efforts. We extend this prior research by exploring how to analyze provenance and provide feedback for crowdsourced exploratory sensemaking.

## 3 CrowdTrace

### 3.1 Problem Definition

**Crowdsourced Sensemaking.** We focus on refining crowdsourced sensemaking that encompasses the holistic process described in the sensemaking loop [22] (Fig. 1). Taking mystery solving as an example, the problem came with *source data*, which is a set of text documents. The crowds collaborated through different stages to solve the mystery from the source data. *Step 1* selects the relevant documents. *Step 2* extracts important information from the relevant documents. *Step 3* identifies possible answers and tags the corresponding supporting evidence. *Step 4* ranks the likelihood of all possible answers. *Step 5* expounds on how the best answer fits into the known facts. Each step is crowdsourced and the crowd results of each step are passed to the next step as input [16]. We refer to the outputs of all the steps as *crowd data*. We visualize both the source data and crowd data to support refinement of the crowd data and make progress on solving the mystery.

**"Problems" in Crowd Analysis.** Crowds make mistakes in sensemaking, just like experts. The challenge is, in exploratory analysis such as solving mysteries, it is hard to define what a mistake looks like without knowing the correct answer. In order to evaluate the crowd work and to steer the refinement, we define the goal of refining crowd analysis to be diagnosing and fixing the problems in the crowd analysis. Similar to the top-down processes in the sensemaking loop [22], a problem can be identified through inconsistencies in the crowd analyses. For example, if Doc A was considered as irrelevant but Doc B contains clues that reveal Doc A's relevance, then overlooking Doc A and the evidence it contains could be a problem. Re-examination of the collected evidence in a broader context could reveal inconsistencies in distributed analysis, or suggest new patterns for alternative hypotheses.

**Crowd Auditing.** CrowdIA [16] established a crowdsourced sensemaking pipeline that facilitates crowd collaboration in distributed sensemaking. A follow-up study [17] evaluated this pipeline

and developed a typology of crowd errors in distributed sensemaking. These prior works illustrated the potential of novice crowds in complex sensemaking, but also pointed to the challenge of iterative refinement of the crowd work. In this work, we explore how to improve a given set of distributed sensemaking analyses.

To support the refinement of pipelined crowd analysis, we conducted a series of preliminary studies to explore 1) who can refine the analysis, and 2) what are the sub-tasks in the refining process. After experimenting in different settings with crowds and individual lab participants, we began to see the emerging role of the committed analyst as a kind of *auditor*. In the business world, auditors are external analysts with two key responsibilities: 1) finding problems within an organization, and 2) proposing solutions. Taking inspiration from this model, we conceptualize the analyst's goal as *crowd auditing*, which is to trace and identify the problems in crowd analysis, and provide actionable feedback to fix the problems.

Based on pilot studies and prior work [20], we identified the following 4 tasks in crowd auditing. **T1: Analysis status overview.** Auditors need to first understand the original workflow through which the crowd contributed the analyses. Similar to the sensemaking loop, there are multiple connected sub-processes in which different groups of crowds collaborated. **T2: Trace data transformation.** The crowd contributed "local" analyses of partial data in each sub-process. Auditors need to make sense of what data was presented to each crowd worker, and how the resulting analysis is combined and re-distributed among the next group. **T3: Identify problems in analysis.** Through a systematic review of crowd provenance and tracking the evidence trails, auditors probe inconsistencies in different parts of the analysis and compare the crowd hypotheses with the known facts to discover problems with the analysis. **T4: Formulate feedback.** After identifying the problems, auditors provide feedback and suggest potential local fixes as well as broader process improvements, to steer the refinement of the analysis.

### 3.2 Using the Auditing Interface to Identify Problems

The main component of CrowdTrace is the auditing interface. It displays the global context of the mystery (Figure 1 A) and the crowd analysis provenance (B). The crowd analyses are laid out in the order of provenance in different columns (**T1**). The first column is the raw dataset, and the last column is the final presentation of the crowd analysis. Middle columns are the intermediate step outputs. Each column has multiple data items, such as documents, information pieces, and candidate answers. Clicking on the item titles in each column can expand or collapse the corresponding content.

**Trace the Analysis Provenance** We provide two tracing mechanisms to help auditors understand the crowd workflow (**T1**) and the history of data transformation (**T2**). First, auditors can hover over an item to see the source information and the downstream analyses (C). The arrows help the auditor understand the distribution of data in each step, as well as the local context available to each crowd worker. Second, the auditor can lock the provenance flow by clicking on the Trace button to keep the related items highlighted.

**Search for Keywords and Threads of Evidence** CrowdTrace also allows auditors to search for occurrences of different words and phrases (D). All matched occurrences are displayed by expanding the corresponding items. Other items without any occurrence will be collapsed accordingly. For example, when the auditor searched for "Matthew Hamilton," then Doc 38, Info 89, Profile 170 are expanded because there are matched occurrences. CrowdTrace supports easy search of keywords in the global context by clicking on them directly. With the crowd analyses displayed in order of provenance, coupled with the tracing features, auditors can examine the keywords occurrences in each step, and compare the analysis about the same keyword in different locations to uncover inconsistencies and identify the problems in the analysis (**T3**).

**Annotate Problems and Take Notes** The auditors can highlight the problematic parts and describe the problem (**T3**). Annotations created by the user can be accessed and retrieved through a dropdown list on the upper right. Auditors can review their auditing outcome and refer back to the local context of each annotation.

### 3.3 Providing Feedback by Creating Microtasks

CrowdTrace enables auditors to provide feedback by directly formulating the feedback as microtasks (**T4**). Our preliminary studies suggested that this feature eliminates communication overhead between the auditor and crowd workers, and can help auditors provide more understandable, actionable feedback. When the auditor creates a microtask, CrowdTrace displays a preview of the microtask interface which crowd workers will see. The preview includes a half-completed instruction about task background. Auditors must fill in 4 blanks in the template: 1) describe the problem, 2) provide instructions for the crowds to fix the problem, 3) provide the corresponding input information for the crowds to work on, and 4) specify the requirements of the format of the crowd answers. Auditors can choose to create a microtask from an annotation to carry over the previous insights. The comments in the annotation will be imported to 2) problem description, and the highlighted text will be automatically imported in 3) input information. We also vertically stacked the microtask creation interface on top of the auditing view to facilitate easier referring back to the crowd analysis.

## 4 Evaluation and Results

We conducted a user study to evaluate our system design and inform the design of future tools for supporting crowd auditing.

### 4.1 Dataset and Participants

We evaluated CrowdTrace under the scenario of refining the crowd analysis in a mystery solving process.

**Source Data** The mystery was adapted from a real-world training exercise for intelligence analysts. The dataset was challenging for crowds [16, 17] and requires multiple iterations and hours from committed analysts [2, 6, 25]. There are 15 raw text documents: 10 are relevant to the mystery and 5 are misleading noise.

**Crowd Data** The crowd analysis was generated by crowd workers on Amazon Mechanical Turk (MTurk), guided by the CrowdIA software [16]. First, 45 crowd workers were hired to evaluate the relevance of the 15 raw documents. Each document was assigned to 3 crowd workers and the relevance was decided by majority vote. Next, another group of crowd workers was hired to extract the key information pieces from the relevant documents. Continuing the five steps described in section 3.1, 49 crowd workers participated in the mystery solving and generated the crowd data. For publication purposes, we have obfuscated the names in the dataset.

We recruited 19 lab participants as auditors to evaluate CrowdTrace. The participants were aged 18–29; 5 were female and 14 were male. The participants had no prior experience with crowdsourcing, and were given an hour to complete the study.

### 4.2 Performance Metrics

We evaluate participants' performance by 1) the problems identified and 2) the quality of the microtasks created by each participant. Two of the authors compared the crowd data to the solution of the mystery and developed a list of important problems. *Important problems* are those that prevent the crowds from achieving the correct answer. For example, it is an important problem if the crowds missed a relevant document, whereas forgetting to capitalize the word "USA" is a trivial problem. Each of the two authors first developed a list of important problems separately. Then the two authors consolidated their lists through an in-depth discussion. In total, we identified 26 important problems with the crowd data.

To measure the quality of the auditors' microtasks, we employed a task design and bonus policy inspired by prior work [19]. We

invited crowd workers to work on the microtasks created by the auditor participants and rate the quality of the microtasks.

### 4.3 Effective Problem Identification in Crowd Auditing

The participants created an average of 14 annotations (min=6, max=21, median=12). Two of the authors conducted qualitative analyses on the annotations. Each author first compared the annotations to the gold standard important problems. If the problem matched the list of important problems, the corresponding ID of that problem was noted. The two authors compared the coding and consolidated the differences in the qualitative analysis of annotations (inter-rater agreement k=0.82, i.e., very good agreement).

Overall, most annotations are about important problems in the analysis (Fig. 2). In addition to the important problems, we also found annotations that describe "other problems" (e.g., typos or grammar errors); "auditor mistakes" (the comments contain a mistake, e.g., considering a correct analysis as wrong); or "note to self" (the comments do not identify a problem or contain any mistakes).

Meanwhile, all the important problems were identified by at least one participant. While most problems (15 out of 26) were identified by 6 or more participants, some problems were more successfully identified than others. For example, all 19 participants identified the problem of an alias of a terrorist being missing. On the other hand, only one participant discovered that a piece of extracted information about a particular terrorist was missing in the final presentation. However, the success of problem identification did not show a clear relation to the difficulty of the problems. For example, Problems 1, 2, and 3 all occurred in the first step of the crowd pipeline and had roughly the same difficulty (all involved irrelevant documents being included in the crowd analysis). However, Problem 1 was identified by 13 participants, problem 2 was identified by 12, but problem 3 was identified by only 5 participants.

We asked the participants about their experience working on the auditing tasks and using CrowdTrace. Almost all participants (N=17) found CrowdTrace helpful for crowd auditing tasks (**T1, T2, T3, T4**). Many participants (N=11) mentioned that being able to *trace* the provenance helped them understand the given analysis (**T1, T2**). P4 said that tracing *"helped me easily visualize where multiple information pieces were coming from in regards to a person analysis."* P15 found tracing helpful for *"seeing where the transition of information broke down and writing a task on it"* (**T3, T4**). The simulated view of microtasks helped participants to communicate their feedback to crowds (**T4**): *"Creating tasks was very easy to use, especially once the annotation was made"* (P13).

### 4.4 Actionable and Clear Microtask Creation

The participants created an average of 6 microtasks each, for a total of 110 microtasks. We created a HIT (Human Intelligence Task) on MTurk for each microtask and hired 3 crowd workers to evaluate each microtask (330 workers hired in total). On average, each crowd worker spent 6.5 minutes and was paid $0.96. 254 (77%) workers rated the problem specification in the given microtask as "very clear" or "clear." 246 (76%) workers rated the task specification in the given microtask as "very clear" or "clear". 244 (74%) workers rated the input information in the given microtask as "very sufficient" or "sufficient." Finally, 237 (72%) workers rated the format requirement in the given microtask as "very clear" or "clear."

## 5 Discussion and Conclusion

In this paper, we contribute the concept of *crowd auditing*, a way to diagnose problems in distributed sensemaking. While crowd workers apply local contexts to generate an analysis, auditors apply a global context to review the analysis. We developed CrowdTrace, a crowd auditing tool to demonstrate the feasibility of the concept.

**Due diligence: awareness of global context**. In CrowdTrace, we found that having always-visible known clues and highlighting

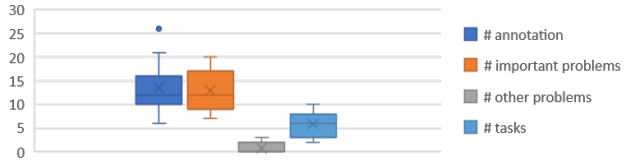


Figure 2: The number of annotation and tasks created by participants, and the number of important problems identified.

the keywords helped auditors focus on the overall analysis goal. Auditors who identified more important problems, such as P13 and P18, focused more on the known clues and followed the lead in the crowd analysis. Less successful auditors tended to be distracted by compounded errors. For example, P16 spent most of his time trying to figure out how an irrelevant person was related to the attack. Future work can explore how to better assist auditors in focusing on the global context, such as by visualizing the analysis about each known clue (dimension coverage [27]). This principle echos the *due diligence* auditing process in business contexts [11]. The goal is to evaluate the "climate" of the business and establish the objectives and postulates to plan and scope the later auditing effort.

**Analytical procedures: awareness of local context**. Crowdsourced analyses are usually constructed from local analyses of different textitcontext slices [18]. A crowdsourcing novice may wonder "why is this information not used, while being available?" Understanding the available local context for previous crowds (**T1** or *perceive* in [20]) is important to identify problems and provide feedback. Furthermore, tracing the analysis provenance (**T2**) also enables auditors to focus on one thread of clues at a time and mitigate information overload. This principle echos auditing techniques such as inspection and inquiry in financial auditing [9]. The goal is to establish an understanding of the specific client processes, evaluate the information available, and probe relationships among the data.

**Re-slice the context for refinement**. Refining the analysis often requires compensating for the missing context in the previous distribution of crowd work. Keyword search with auto expand/collapse enables auditors to organize the data and analyses into new context slices to support the effective refinement of the current analyses. For example, all 19 participants identified the problem of an alias (Michael) of a terrorist (Hugh Ash) being missing. They searched for the keyword *Hugh Ash* and discovered the name in two documents that were split into different context slices. The participants created a new context slice for these two documents and requested more information about Hugh and Michael. This principle echos the selection and sampling techniques [8] for obtaining audit evidence and drawing implications for audit reports in financial accounting.

This work utilized individual auditors. Future research is needed to explore how multiple auditors can collaborate in crowd auditing. Such collaboration among domain experts can benefit from prior work on collaborative sensemaking [27, 33]. However, collaboration among transient novice crowd workers poses unique challenges in distributing the auditing context and progress.

In conclusion, visualizing the sensemaking provenance can help debug and produce better analyses. Crowdsourced pipelines (like CrowdIA) serve to formalize the sensemaking process, which then affords visualization of provenance (like CrowdTrace). In the future, the concepts and lessons of crowd auditing could be more broadly applied to individual or collaborative scenarios to improve analysis and enable meta-level sensemaking.

## Acknowledgments

This research was supported in part by NSF grants IIS-1527453, IIS-1651969, and IIS-1447416.

## REFERENCES


[1] P. André, A. Kittur, and S. P. Dow. Crowd synthesis: Extracting categories and clusters from complex data. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, p. 989–998. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2531602.2531653

[2] J. Cheng, R. Emami, L. Kerschberg, Q. Zhao, H. Nguyen, H. Wang, M. Huhns, M. Valtorta, J. Dang, H. Goradia, et al. Omniseer: a cognitive framework for user modeling, reuse of prior and tacit knowledge, and collaborative knowledge services. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pp. 293c–293c. IEEE, 2005.

[3] G. Chin, O. A. Kuchar, and K. E. Wolf. Exploring the analytical processes of intelligence analysts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 11–20. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518704

[4] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics and Applications*, 29(3):52–61, 2009.

[5] S. Dow, A. Kulkarni, S. Klemmer, and B. Hartmann. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, p. 1013–1022. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2145204.2145355

[6] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 473–482, 2012.

[7] K. Fisher, S. Counts, and A. Kittur. Distributed sensemaking: Improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 247–256. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2207711

[8] R. Florea and R. Florea. Audit techniques and audit evidence. *Economy Transdisciplinarity Cognition*, 14(1), 2011.

[9] S. Glover. Analytical procedures and audit planning decisions: Unexpected fluctuations that influence auditors to revise their audit plans. *Journal of Accountancy*, 191:99, 02 2001.

[10] N. Hahn, J. Chang, J. E. Kim, and A. Kittur. The knowledge accelerator: Big picture thinking in small pieces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, p. 2258–2270. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858364

[11] M. G. Harvey and R. F. Lusch. Expanding the nature and scope of due diligence. *Journal of Business Venturing*, 10(1):5–21, 1995.

[12] A. Kittur, A. M. Peters, A. Diriye, T. Telang, and M. R. Bove. Costs and benefits of structured information foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 2989–2998. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2481415

[13] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent systems*, 21(5):88–92, 2006.

[14] G. Klein, J. K. Phillips, E. L. Rall, and D. A. Peluso. A data–frame theory of sensemaking. In *Expertise out of context*, pp. 118–160. Psychology Press, 2007.

[15] T. Kulesza, S. Amershi, R. Caruana, D. Fisher, and D. Charles. Structured labeling for facilitating concept evolution in machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 3075–3084. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557238

[16] T. Li, K. Luther, and C. North. Crowdia: Solving mysteries with crowdsourced sensemaking. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), Nov. 2018. doi: 10.1145/3274374

[17] T. Li, C. J. Manns, C. North, and K. Luther. Dropping the baton? understanding errors and bottlenecks in a crowdsourced sensemaking pipeline. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), Nov. 2019. doi: 10.1145/3359238

[18] T. Li, A. Shah, K. Luther, and C. North. Crowdsourcing intelligence analysis with context slices. In *Chi'18 Sensemaking Workshop*, 2018.

[19] V. C. Manam and A. J. Quinn. Wingit: Efficient refinement of unclear task instructions. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*, 2018.

[20] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: Process+interaction+insight. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, p. 33–36. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1979742.1979570

[21] W. A. Pike, J. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, Dec. 2009. doi: 10.1057/ivs.2009.22

[22] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, vol. 5, pp. 2–4. McLean, VA, USA, 2005.

[23] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):31–40, 2016.

[24] D. Retelny, M. S. Bernstein, and M. A. Valentine. No workflow can ever be enough: How crowdsourcing workflows constrain complex work. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–23, 2017.

[25] A. C. Robinson. Collaborative synthesis of visual analytic results. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pp. 67–74. IEEE, 2008.

[26] J. Rzeszotarski and A. Kittur. Crowdscape: interactively visualizing user behavior and output. *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pp. 55–62, 2012.

[27] A. Sarvghad and M. Tory. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the 41st Graphics Interface Conference*, p. 123–130. Canadian Information Processing Society, CAN, 2015.

[28] N. Sharma. Sensemaking handoff: Theory and recommendations. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, p. 1673–1676. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1240866.1240880

[29] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 1237–1246. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357247

[30] A. Toniolo, T. J. Norman, A. Etuk, F. Cerutti, R. W. Ouyang, M. Srivastava, N. Oren, T. Dropps, J. A. Allen, and P. Sullivan. Supporting reasoning with different types of evidence in intelligence analysis. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, p. 781–789. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2015.

[31] W. Willett, S. Ginosar, A. Steinitz, B. Hartmann, and M. Agrawala. Identifying redundancy and exposing provenance in crowdsourced data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2198–2206, 2013.

[32] S. Xu, C. Bryan, J. K. Li, J. Zhao, and K.-L. Ma. Chart constellations: Effective chart summarization for collaborative and multi-user analyses. *Computer Graphics Forum*, 37(3):75–86, 2018. doi: 10.1111/cgf.13402

[33] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE transactions on visualization and computer graphics*, 24(1):340–350, 2017.