

Analytic Provenance and Distributed Sensemaking

Ashley Wheat, Simon Attfield, and Robert Fields

Abstract Analytic provenance is a record of reasoning over time, accounting for the methods and techniques used. In sensemaking—where people embark on a process of comprehension by which they gain meaning and insight from information—a record of provenance can support the scrutiny of findings, reflection on the reasoning process, and handover of tasks in collaborative settings. However, sensemaking does not occur within a vacuum, and often involves use of various representational media and artifacts such as maps, charts and lists to gain insight. Therefore, a complete account of analytic provenance in sensemaking scenarios must include descriptions of the use of these representational media. In this paper we discuss analytic provenance in the context of *distributed sensemaking*, showing how we can model the use of representational artifacts and reasoning over time as *inference trajectories*, introduce levels of description of representational artifacts and discuss challenges faced in the capture of analytic provenance in distributed sensemaking scenarios.

1 Introduction

Sensemaking refers to a process of comprehension by which human beings formulate a plausible understanding and explanation from information we receive from the world around us. When carrying out complex sensemaking tasks it can be important to maintain a record of the reasoning process. In contexts such as law and intelligence analysis it is imperative that a ‘chain of custody’ or ‘paper trail’ is maintained, keeping track of the control and analysis of data and information. This historical account of an analysis can help reduce uncertainty and increase trust in findings by allowing reasoning to be scrutinized, supports handover of analysis in collaborative settings, and can support the sensemaker’s own understanding and confidence in their analysis. This historical account of reasoning in an analysis, known as its ‘analytical provenance’ [1], provides a description of the actions performed and techniques used at a given point in an analysis.

A. Wheat (✉) • S. Attfield • R. Fields
Interaction Design Centre, Middlesex University, London, UK
e-mail: a.wheat@mdx.ac.uk; s.attfield@mdx.ac.uk; b.fields@mdx.ac.uk

Although we have the ability to easily record actions and events in computer environments that can form part of an account of analytic provenance, this only paints part of the picture. An analysis does not take place in a vacuum, and sensemaking does not just take place in a person's head, but through elicitation and interaction with various artifacts and forms of representational media. Therefore an account of analytical provenance must include descriptions of the use and role of representational media in the sensemaking process leading to insights and findings.

In this paper we introduce the notion of *distributed sensemaking* and discuss how its concepts can help in creating a record of analytical provenance that includes an account of the role of representational media in sensemaking. Distributed sensemaking models the flow of information in and co-ordination of representational artifacts to form insights in sensemaking as 'inference trajectories', and provides a number of levels of description in characterizing representational artifacts.

The rest of this paper is structured as follows: in the next section we introduce distributed sensemaking, outlining its theoretical background before introducing the notion of inference trajectories and a number of levels of description. In Sect. 3 we discuss the concept of analytic provenance and the challenges in capturing a full account of analytic provenance including the use and role of representational media. We then go on to discuss how the concepts of distributed sensemaking can provide a foundation for research into the modeling and framing of analytic provenance in terms of the use of representational media and artifacts in sensemaking. Lastly we discuss challenges faced in capturing this type of provenance information across numerous types of media.

2 Distributed Sensemaking

2.1 Sensemaking

The term sensemaking literally refers to 'making sense'. For instance, sensemaking can take place when a holiday maker is comparing the best flight deals online, or when a detective is examining evidence in order to find the culprit responsible for committing a crime. When we engage in sensemaking, we embark on a process of comprehension [2] in which we seek out, re-structure and re-organize information in order to find meaning and construct a plausible understanding of some aspect of the world [3, 4]. Multiple theories of sensemaking have emerged—seemingly independently—in a number of research areas [5] including Organizational Studies [3], Information Science [6], Human-computer Interaction [7, 8] and Naturalistic Decision Making [2, 9].

Klein et al. [2, 9] offer a 'macrocognitive' theory of sensemaking involving the interaction of two types of entity: *data* and *frames*. Data are aspects of the world as experienced by the sensemaker through interaction with it. These might include things that a person might perceive in a given situation that may be important to

them, such as a patient's symptoms in a medical setting or the co-ordinates and direction of aircraft in air traffic control.

A frame is a representation that accounts for the current understanding of something. For example, this could be the doctor's belief about the patient's medical condition, or it could be the air traffic controller's understanding of the flightpaths of aircraft in airspace he or she is responsible for. In this light, a frame serves as both an interpretation and explanation of the data available at a particular moment in time [10]. According to Klein et al., sensemaking is a continual process involving framing and re-framing when new data is available. As the sensemaker experiences a new situation, a frame acts as an interpretation of it. As more data becomes available, the current frame may be elaborated upon or challenged, causing the frame to evolve over time. As it does, it becomes a more plausible account of the situation as previous frames are rejected or modified in light of new data. Furthermore, as sensemaking is a bi-directional process, a frame may also call upon new data to be sought out, directing information seeking, and in so doing, revealing further data that changes the frame.

2.2 *Distributed Cognition*

Distributed cognition provides a perspective in which human cognition transcends the boundaries of the head of the individual, seeing intelligent processes as being distributed among people, the artifacts they use and the environment in which they are situated, and is affected by previous events and experiences [11].

In distributed cognition, cognitive activities are seen as computations that propagate representational state through a series of different media, which can occur both inside or outside of the head. For example this could be a person's memory, or external media such as charts or maps. The unit of analysis in distributed cognition, therefore, is a cognitive system which is made up of the internal processes of individuals interacting with a number of artifacts, each other and the environment in which they are situated. Studies of such cognitive systems include ship navigation [11], aircraft cockpits [12, 13], air traffic control [14] and emergency medical dispatch [15].

Hollan and colleagues [16] describe distributed cognition as three 'tenets': *socially distributed cognition*, which describes the distribution of cognitive tasks among individuals within a social group; *embodied cognition*, which describes the coordination between internal (the mind) and external (materials and environment) functions; and *culture and cognition*, describing how cognitive processes can be shaped by earlier experiences or social and cultural practices.

2.3 Representational Media in Sensemaking

Embodied within artifacts and representational media (e.g. maps, charts, lists) are a number of affordances which can furnish people with the ability to perform tasks that may otherwise be difficult to conduct solely in the head. These representations occurring ‘in the world’ are thought to change sensemaking in some way, but aren’t addressed in much depth in existing sensemaking theory [10]. *Distributed sensemaking* addresses this by considering sensemaking through the lens of distributed cognition.

2.4 Inference Trajectories

In the distributed sensemaking paradigm, the flow of information throughout the sensemaking process across different representational media can be modeled as *inference trajectories*. An inference trajectory shows the relationship between information about some aspect of the world, extracted from representational artifacts, and its use in conjunction with information contained within other representational media (which could be internal or external). When used in conjunction with each other, these pieces of information (and the media they are contained within) lead to the generation of insights and a *situation picture*. A situation picture, similar to Klein’s frame [9], is a sensemaker’s current understanding of a given situation, representing a plausible picture of events taking place in the real world.

A situation picture can be represented either internally (in the head) or externally (in the world), embodied within some representational artifact. As a sensemaker gains more traction in their reasoning process, gaining more insight and understanding, the situation picture becomes clearer and more well defined.

Figure 1 illustrates an inference trajectory from the study of military signals intelligence analysis. The study was conducted on analysts within a military signals

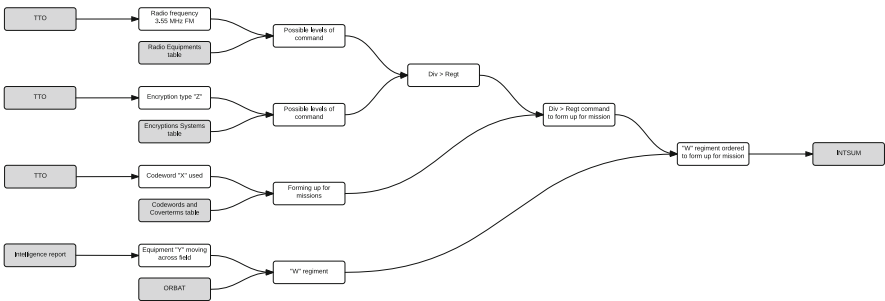


Fig. 1 An illustration of an inference trajectory showing the relationship and coordination of representational media in military signals intelligence analysis. Physical artifacts such as tables and charts are shown in gray. White nodes show information

intelligence cell, who's job it is to gain an understanding of the identity of enemy assets, their level in the command structure, their equipment and movements. In the study, analysts were fed extracts of intercepted radio communications in the form of 'tactical tip off' reports, or TTOs, using a number of 'working aids' to perform their analysis. These consisted of tables and charts containing known information about the opposing force, such as radio equipment information, known call signs, known use of code words, and intelligence about the command hierarchies and formations known as an Order of Battle (ORBAT). In the top-left (Fig. 1) the inference trajectory shows information extracted from a TTO—a radio frequency of 3.55 MHz FM—used in conjunction with a 'Radio Equipment table' leading to a number of possible levels of command. A similar operation is performed by the analyst revealing a further number of possible levels of command when the analyst uses information extracted from a TTO about an enemy asset using a radio encryption in conjunction with the 'Encryptions Systems table'. From this the analyst infers that the enemy level of command is 'Div → Regt'. This inference comes as the result of a boolean conjunction between the two lists of possible levels of command (for further explanation see [10]) and would be difficult to perform without the use of external representational media—in this case two tables of information.

2.5 Levels of Description in Distributed Sensemaking

Inference trajectories provide an abstract view of the information flow and coordination of representational media within the sensemaking process. However, the properties of artifacts leveraged by the sensemaker are also key to their use. For example, in the study described above, the analyst used a number of tables to carry out the sensemaking task. One such table was the 'Radio Equipments table', which contained known information about enemy radio equipment including frequency ranges, modes (FM, AM etc.) and levels of command within the military hierarchy which use certain frequencies. When working out possible levels of command of intercepted communications, the analyst would refer to the table and eliminate row by row—by striking through using a pen or pencil—those radio frequency ranges and types which the intercepted signal do not match. By doing this the analyst is deductively working out a list of possible levels of command. Moreover, as the analyst strikes out each row on the table, he reduces the number of possible levels of command for a signal, leading to an clearer situation picture.

We describe such properties at three distinct levels of description: *physical*, *semantic* and *pragmatic*.

Physical properties can be described in terms of an artifact's material and shape—how it is physically constituted. Moreover, when considering the physical makeup of an artifact, the affordances it offers in virtue of them are also considered. That is,

the physical properties of an object which help, support, facilitate or enable physical action [17].

Semantic properties of artifacts are what they are taken to represent or stand for. That is, when used in sensemaking, artifacts are imbued with some representational meaning such that they represent some aspect of the world. For example, a database in a shop might represent a series of associations between products/stock levels and cost.

Pragmatic properties like the semantic properties of artifacts are concerned with their meaning and what they represent. However, where the semantic meaning of an artifact is constant, the pragmatic properties of artifacts are concerned with the role given to the artifact in current cognitive activity, which is subject to change. Namely, this is what an object is used for in virtue of its physical and semantic properties. For example, a shopping list might have items crossed or ticked off as they are put in the shopping trolley. To the shopper, this represents a list of items retrieved (crossed or ticked off) and items needed (not crossed or ticked off). Each time the shopper crosses off an item retrieved, the shopping list gains new meaning in terms of its cognitive role—it serves as an up-to-date record of items in the shopping trolley, and items not yet collected.

3 Analytic Provenance and Distributed Sensemaking

3.1 Analytic Provenance

An account of analytic provenance can be important in many situations, helping to reduce uncertainty and aid collaboration. Analytic provenance accounts for the actions and techniques used in an analysis at any point in time. In areas such as legal practice for example, it is important that a ‘chain of custody’ or ‘paper trail’ is preserved showing the control, transfer and analysis of evidence. By maintaining a record of analytic provenance, an account of the analytic process at any point is kept. This supports “reflection-on-action” [18] by allowing the interpretation and audit of claims and insights to be made, preserving a level of accountability and confidence in findings. Provenance information can also be important during an analysis itself by supporting “reflection-in-action”, allowing people to interpret their own findings, identify areas in their analysis that might be weak and help them make sense of what they are trying to do [19]. Furthermore, in collaborative contexts—where analysts may be working as part of a large team or in non-colocated settings—an account of provenance can play a vital role in keeping track of individual actions which may not be clear from results alone [19]. This can be useful in assisting the coordination of labour, enabling best practice and supporting handover of tasks in an analysis.

3.2 Provenance and Representational Media

When sensemaking occurs, in many contexts reasoning takes place through the elicitation of a number of resources and representational media, both internally and externally. In Sect. 2 we discussed how we view this as distributed sensemaking. In light of that, it must be taken into consideration when recording analytic provenance, that there may be a variety of different sources of insight and knowledge. We previously introduced the study of military analysis which used a number of printed charts and tables known as ‘working aids’ alongside computer software and tools to generate insights and knowledge. The use of this type of representational media and external resources is commonplace, and as such, any account of analytic provenance may be seen as incomplete without a record of the flow of information and inference generation through the use of representational artifacts.

4 Challenges in Framing and Capturing Analytic Provenance

According to Xu et al. [19] there has been considerable progress in the capture and visualization of data provenance and analytic provenance, however, there is still some progress that can be made until it can be understood and used in terms of distributed sensemaking. Currently there is the ability within visual analytics systems to capture events such as mouse clicks, keystrokes and actions such as database queries and searches within computer environments [1, 19]. But this provides only part of the picture. As we have discussed, sensemaking occurs through the elicitation of different media, therefore to provide a full account of analytic provenance, we must find ways to capture it across the different representational media used within the sensemaking process. This presents a number of challenges. Firstly, the modeling and framing of this information requires further research to be carried out to have a more complete understanding of distributed sensemaking and to further develop a framework for its capture and analysis. Secondly, given the nature of different representations and artifacts—outside of the computer environment or inside the head—it is very difficult, if not impossible to automatically capture this information.

4.1 Modelling and Framing Analytic Provenance in Distributed Sensemaking

In Sect. 2 we introduced a model of distributed sensemaking including inference trajectories and a number of levels of description of representational artifacts. We believe this provides a foundation for research into the modeling and framing of analytic provenance in distributed sensemaking.

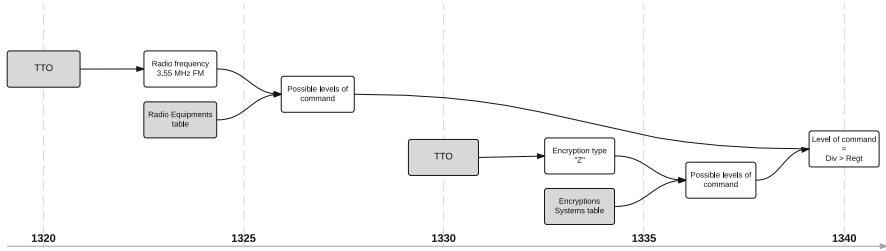


Fig. 2 A section of an inference trajectory showing the use of representational artifacts through time

4.1.1 Inference Trajectories

Inference trajectories show the relationship and coordination between information extracted from, and the use of different representational media in sensemaking scenarios. In Fig. 1 we have shown an illustration of an inference trajectory in military intelligence analysis. This provides a useful bird’s eye view of an analyst’s sensemaking (or analytic) process, showing how he or she has reached inferences and insights through the elicitation of different representational media. However, it does not provide a chronological account showing the development of the sensemaking process and flow of information over time—the analytic provenance. Figure 2 shows a section of the same inference trajectory which has been adapted to reveal the use of representational artifacts and generation of insights through time. By looking at a certain point in the chronology, a snapshot of the distributed sensemaking process—and the representational media involved—can be seen. Moreover, by viewing analytic provenance and the use of representation in this way, an account of events leading to insights and inference is visible in order of occurrence, allowing easy reflection on the analytic process and the status of information and knowledge at an point in time.

4.1.2 Levels of Description

Inference trajectories are a useful way of looking at the overall use of representational media and artifacts. However, this comes at a low level of resolution, and provides no detail about the make-up of artifacts or details of how they are used, which is important when reflecting on the use of representational media in an analysis or sensemaking task. In Sect. 2.5 we introduced a number of levels of description within the distributed sensemaking paradigm. These look at the *physical properties* of an artifact such as its material and shape and the physical affordances it offers; the *semantic properties* of an artifact, which look at the representational meaning given to an artifact; and the *pragmatic properties* of an artifact, which look at its role in current cognitive activity.

So, by describing the properties of representational artifacts at these levels throughout given points in the inference trajectory, we can show the use of representational artifacts at a given point in an analysis as well as how their use lead to insight and knowledge generation.

4.2 Capturing Analytic Provenance in Distributed Sensemaking

The capture of analytic provenance is a significant challenge. The capture of low level events and actions in digital environments is relatively easy [19]. However, this type of provenance information reveals only a limited picture of the sensemaking process, as much of this occurs outside of the computer environment across different physical media and inside the sensemaker’s mind. The capture of analytic provenance therefore must occur, in part, manually. This however can be time consuming and labor intensive. Another issue is that of timeliness. It may be that the sensemaker may forget what they were doing at a given point, or what their thinking was when using a representational artifact, so without capturing provenance information within a limited timeframe, it could be lost or become less reliable.

There are contexts where analytic provenance—across different media—is already captured and forms an important part of maintaining reliability and trust in information. We previously mentioned this in the context of law, where a ‘paper trail’ of evidence must be preserved documenting the acquisition, control and analysis of evidence. In fields such as history, art and archival sciences a similar chronology of the status of artifacts must be maintained to determine authenticity. In these areas, the capture and recording of provenance information is already established, and may prove to be fruitful areas for research when facing the challenge of capturing and recording analytic provenance in distributed sensemaking scenarios. By conducting such research, we could learn efficient and well established methods of acquiring, documenting and preserving provenance information, which could be applied in the capture of analytic provenance.

5 Conclusion

Sensemaking does not occur only in the head of the individual, but through the elicitation of, and interaction with various forms of representational media. Therefore, a full account of analytic provenance in sensemaking scenarios must describe the use and role of the different representational media and representational artifacts in the process. This full account of the sensemaking process over time can be useful in a number of ways. It can support “reflection-on-action” by capturing

points in the sensemaking process where resources are used together in order to reach insights, allowing the scrutiny and validation of findings, thus reducing uncertainty. Moreover, by reflecting on the use of representational media we can learn from the reasoning process, developing better materials and resources for sensemaking and analysis. It can also be a source of “reflection-in-action” allowing individuals to interpret their own findings and identify weaknesses in their analyses, as well as supporting collaboration by keeping track of individual actions.

In this paper we have shown how inference trajectories can keep track of the use of representational media over time, and in Fig. 2, we have illustrated this in a military intelligence analysis scenario. Also, by describing the physical, semantic and pragmatic properties of artifacts used at given points in sensemaking we can show how their use impacts on the sensemaking process.

There remains a number of challenges however. A key challenge is that of capturing analytic provenance in distributed sensemaking. Recording an account of analytic provenance which includes the use of representational artifacts must currently be done manually, which is time consuming and labor intensive. Also, there is a limited timeframe by which this information can be collected—a person may forget what they were doing or what they were thinking when using an artifact after a certain amount of time.

Looking ahead, we propose future research, including the study of distributed sensemaking in a number of scenarios, tracking reasoning and the use of representational media through the construction of inference trajectories. Here we can assess the utility of inference trajectories in these scenarios and further develop ways of modeling and framing analytic provenance in distributed sensemaking. We also propose research be carried out in finding more reliable and less costly methods for recording analytic provenance in distributed sensemaking contexts, for example by electronically tagging and tracking the use of artifacts in the environment, facilitating the automatic capture of their use through time.

References

1. Gotz, D., Zhou, M.X.: Characterizing users' visual analytic activity for insight provenance. *Inf. Vis.* **8**, 42–55 (2009)
2. Klein, G., Phillips, J.K., Rall, E.L., Peluso, D.A.: A data-frame theory of sensemaking. In: *Expertise Out of Context. Proceedings of the Sixth International Conference on Naturalistic Decision Making*, pp. 113–155. Psychology Press, Hove (2007)
3. Weick, K.E.: *Sensemaking in Organizations*, vol. 3. Sage, Beverly Hills, CA (1995)
4. Weick, K.E., Sutcliffe, K.M., Obstfeld, D.: Organizing and the process of sensemaking. *Organ. Sci.* **16**, 409–421 (2005)
5. Blandford, A., Atfield, S.: Interacting with information. *Synth. Lect. Hum. Centered Inform.* **3**, 1–99 (2010)
6. Dervin, B.: An overview of sense-making research: concepts, methods and results to date. In: *International Communication Association Annual Meeting* (1983)

7. Pirolli, P., Card, S.: Information foraging in information access environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '95, New York, pp. 51–58. ACM/Addison-Wesley, New York/Reading, MA (1995)
8. Russell, D.M., Stefik, M.J., Pirolli, P., Card, S.K.: The cost structure of sensemaking. Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93, pp. 269–276 (1993)
9. Klein, G., Moon, B., Hoffman, R.R.: Making sense of sensemaking 2: a macrocognitive model. *IEEE Intell. Syst.* **21**, 88–92 (2006)
10. Attfield, S., Fields, B., Wheat, A., Hutton, R.J.B., Nixon, J., Leggatt, A., Blackford, H.: Distributed sensemaking: a case study of military analysis. In: 12th International Conference on Naturalistic Decision Making (2015)
11. Hutchins, E.: *Cognition in the Wild*. MIT, Cambridge, MA (1995)
12. Hutchins, E.: How a cockpit remembers its speeds. *Cogn. Sci.* **19**, 265–288 (1995)
13. Hutchins, E., Klausen, T.: Distributed cognition in an airline cockpit. In: *Cognition and Communication at Work*, MIT, Cambridge, MA, pp. 15–34 (1996)
14. Halverson, C.A.: *Inside the cognitive workplace: new technology and air traffic control*. Ph.D. thesis, University of California, San Diego (1995)
15. Blandford, A., William Wong, B., Wong, B.L.W.: Situation awareness in emergency medical dispatch. *Int. J. Hum. Comput. Stud.* **61**, 421–452 (2004)
16. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput. Hum. Interact.* **7**, 174–196 (2000)
17. Hartson, R.: Cognitive, physical, sensory, and functional affordances in interaction design. *Behav. Inform. Technol.* **22**, 315–338 (2003)
18. Schon, D.A., DeSanctis, V.: The reflective practitioner: how professionals think in action. *J. Contin. High. Educ.* **34**, 29–30 (1986)
19. Xu, K., Attfield, S., Jankun-Kelly, T.J., Wheat, A., Nguyen, P.H., Selvaraj, N.: Analytic provenance for sensemaking: a research agenda. *IEEE Comput. Graph. Appl.* **35**, 56–64 (2015)