

Methodology for Knowledge Elicitation in Visual Abductive Reasoning Tasks

Michael J. Haass^(✉), Laura E. Matzen, Susan M. Stevens-Adams,
and Allen R. Roach

Sandia National Laboratories, Albuquerque, NM, USA
{mjhaass, lematze, smsteve, raroach}@sandia.gov

Abstract. The potential for bias to affect the results of knowledge elicitation studies is well recognized. Researchers and knowledge engineers attempt to control for bias through careful selection of elicitation and analysis methods. Recently, the development of a wide range of physiological sensors, coupled with fast, portable and inexpensive computing platforms, has added an additional dimension of objective measurement that can reduce bias effects. In the case of an abductive reasoning task, bias can be introduced through design of the stimuli, cues from researchers, or omissions by the experts. We describe a knowledge elicitation methodology robust to various sources of bias, incorporating objective and cross-referenced measurements. The methodology was applied in a study of engineers who use multivariate time series data to diagnose the performance of devices throughout the production lifecycle. For visual reasoning tasks, eye tracking is particularly effective at controlling for biases of omission by providing a record of the subject's attention allocation.

Keywords: Knowledge elicitation · Eye tracking · Abductive reasoning

1 Introduction

Knowledge elicitation, the process of understanding an individual's or group's tacit knowledge so that it can be preserved and disseminated as explicit knowledge, is used in a wide range of research and development fields including psychology, decision analysis and artificial intelligence. Application of knowledge elicitation methods can provide important information about expert judgment and decision making. This information can then be used in the design of expert systems to automate or augment human decision making. The potential for bias to affect the results of knowledge elicitation studies, on the part of both researchers and subjects, is well recognized. Researchers and knowledge engineers commonly attempt to prevent, or at least control for, sources of bias through careful selection of elicitation and analysis methods [1]. However, the development of a wide range of physiological sensors, coupled with fast, portable and inexpensive computing platforms, has added an additional dimension of objective measurement that can reduce effects of unanticipated bias. Incorporating these technologies into knowledge elicitation methodologies can also reinforce the validity of qualitative findings and highlight previously undetected biases.

Abductive reasoning is the process of forming a conclusion that best explains observed facts. This type of reasoning plays an important role in the development and application of expertise in many fields such as scientific research, economics and medicine. A common example of abductive reasoning is medical diagnosis. Given a set of symptoms, a doctor determines a diagnosis that best explains the combination of symptoms. Abductive reasoning has been studied in the fields of logic [2], medicine [3], logistics [4] and artificial intelligence [5], however, little work is available on human visual information processing and reasoning during abductive reasoning tasks. We begin to address this research gap using a knowledge elicitation approach. In the case of an abductive reasoning task, biases during knowledge elicitation can be introduced through design of the stimuli, cues from researchers, or omissions by the experts. Bias introduced through design of the stimuli can arise from the designer's choices for spatial layout of information, or even what information is included or excluded from the stimuli set. Biases of omission, or actual-ideal discrepancies,¹ by the expert may be conscious or unconscious.

This paper describes a methodology for knowledge elicitation from experts for a complex visual abductive reasoning task. The methodology is designed to be robust to various sources of bias by incorporating both objective and cross-referenced measurements. We have applied this methodology in a study of engineers who use multivariate time series data to diagnose the performance of devices throughout the production lifecycle. This study will be used to demonstrate the application of the methodology and to illustrate key findings enabled by this approach.

2 Knowledge Elicitation Methodology

To create a knowledge elicitation method robust to unanticipated biases, we incorporate objective, physiological measures during domain-specific tasks that provide cross referencing information for a verbal walkthrough protocol. We also design the study instruments to include elements that could be cross-referenced during analysis to highlight both consistencies and discrepancies in the raw data. Figure 1 illustrates the process used to design the knowledge elicitation session and instruments.

When knowledge elicitation researchers encounter a new work domain, as was the case with this study, it is advisable to design the elicitation methodology to include multiple opportunities to engage with subject matter experts. Therefore, the first step in the process, which can be omitted if the design team already has a wealth of experience in the work domain, is to interview one subject matter expert (expert point of contact, ePOC) and produce a document describing the work domain and culture (outputs and work products of each step are listed in the right hand column of Fig. 1). This step helps ensure that the elicitation methodology will be compatible with the work culture and the availability of the work domain experts. The second step is to conduct observations of a small set of work domain experts as they perform the work tasks to be studied. The output

¹ We use the terminology "actual-idea discrepancies" to describe instances where there are differences between and expert's verbal description of a work process and the actual process followed during completion of the work task.

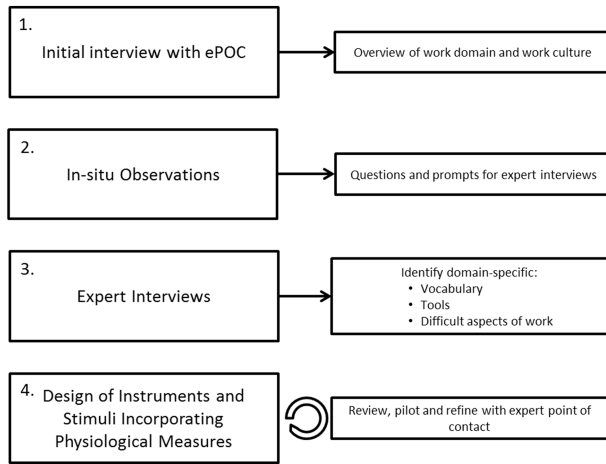


Fig. 1. Process used to design the knowledge elicitation session and instruments.

of this step is a document listing specific questions and more general talking points to facilitate the detailed interviews in Step 3. The third step is detailed interviews with as many work domain experts that time, staffing, or budget will allow. The outputs of this stage are detailed lists of domain-specific vocabulary, tools and difficult aspects of the work. This information is then used in Step 4 to design the instruments and stimuli for the knowledge elicitation sessions. During Step 4, designer should take care to identify opportunities to cross reference information across instruments and incorporate as many cross reference points as possible. At this point, careful consideration is also given to physiological measures that can provide objective measurements of activity during the chosen tasks. Brookings, Wilson & Swain provide one assessment of psychophysiological variables that can be used to assess workload during complex tasks such as air traffic control [6]. Marshall shows that eye metrics (information from pupil size and point-of-gaze) can be used to discriminate between two cognitive states during complex tasks [7]. Matzen, Haass & McNamara suggest eye tracking to be particularly useful for investigating cognitive biases within visual search domains [8]. Figure 2 illustrates the cross referenced methodology incorporating eye tracking as a physiological measure of allocation of attention during a domain-specific task.

3 Physiological Sensors

Modern manufacturing technologies and increasing consumer interest in self-tracking are providing researchers with an increasingly wide range of physiological sensors. For example, consumer-grade devices are now available to monitor and record an individual's physical activity level, sleep patterns and even brain electrical activity.²

² For example, Nike's Fuelband, Fitbit's One, Emotiv's EPOC.

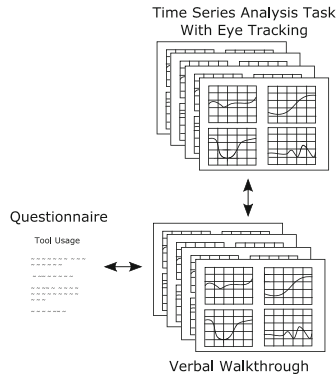


Fig. 2. Cross referenced design: double headed arrows indicate cross referencing between instruments.

Research-grade devices that measure more traditional physiological parameters such as heart and respiration rates have also become more compact and portable and at the same time provide data with increased temporal and measurement resolution.³ Eye tracking is a prime example of these improvements. Modern systems incorporate small, single mount-point camera modules that are easy to transport and can operate from a single USB connection to a laptop computer. These systems record not only gaze point information, but also provide a wealth of information about the subject's head position and orientation. Many systems also provide information derived from eye lid position, such as blinks or eye closures due to fatigue, along with measurement of the pupil diameter for each eye, all at data rates 60 Hz or 120 Hz.

4 Example Application: Visual Abductive Reasoning Study

We applied the methodology in a study of engineers who use multivariate time series data to diagnose the performance of devices throughout the production lifecycle. The goal of the study was to understand the expert's abductive reasoning processes and the key features of the time series data used in these processes. This information can then be used to create and select input features used by advanced data analytics to model and predict certain response variables.

In the work domain studied, access to experts was limited due to their senior roles spanning multiple engineering teams. Therefore, knowledge elicitation sessions had to be as brief as possible, while still being thorough enough to acquire all data relevant to the work and the expert's reasoning processes.

To design the knowledge elicitation sessions, we followed the multi-stage approach described in Sect. 2. In Step 1, we collaborated extensively with the ePOC to gain an overview of the entire work domain and to develop an elicitation methodology that

³ For example, Seeing Machine's FOVIO, BIOPAC's BioHarness, ANT Neuro's eegoSports.

would be compatible with the work culture and the availability of the expert engineers. In Step 2 we conducted observations of a subset of experienced engineers performing their day-to-day time series analysis work. From these observation sessions, we developed a list of specific questions and general talking points to use in subsequent interviews. In Step 3, we conducted one hour interviews with each of the three most experienced engineers. After completing the interviews, we analyzed our notes and developed lists of commonly used, domain-specific vocabulary, tools used to complete the time series analysis work, and common difficulties encountered during the analysis work. In Step 4, we used the information from Step 3 to design the instruments to be used in the knowledge elicitation study. We chose to include eye tracking as an objective measure of attention allocation during a domain-specific task. We also chose to create study instruments that could be cross-referenced during analysis to highlight both consistencies and discrepancies in the raw data. This design allows for comparisons across objective measures of attention and subject measures of information collection and reasoning processes that will help guide subsequent studies in this work domain.

Four instruments were developed, (1) a general demographics questionnaire, (2) a work domain-specific questionnaire, (3) a simplified, domain-specific, abductive reasoning task, equipped with eye tracking, and (4) a verbal walkthrough protocol (see Sect. 4.2 for a detailed description of each instrument). For each instrument, we created an initial draft and then reviewed and revised the content and instructions in collaboration with the ePOC. The ePOC also provided technical content (time series) to use as stimuli for the time series analysis task.

4.1 Participants

Thirteen employees at Sandia National Laboratories volunteered to participate in the study. Three of the participants in the study were classified as experts; that is, they diagnosed device performance using the multivariate time series data as part of their daily job. These experts had an average of 15.5 years' experience performing this type of activity. Four participants were categorized as practitioners; that is, they were familiar with the multivariate time series data but did not use it to diagnose device performance. These practitioners had an average of 5.5 years' experience interacting with the multivariate time series data. Six participants were classified as novices who had no experience with the multivariate time series data. This novice cohort was included to provide comparative performance baselines.

4.2 Procedure

The participants completed the study individually. The participant first read through and signed the study consent form and asked any questions he/she had about the study. Next, the participant filled out a demographic questionnaire which assessed the participant's age, gender, years of experience, etc. The experts and practitioners then filled out a questionnaire which asked specific questions about their work with the

Multivariate time series data. The novices did not fill out this second questionnaire since they did not have any experience with this type of data.

Multivariate Time Series Task. A PowerPoint presentation was displayed to the participant which explained the study and described what the participant would be asked to do. The novices were given very detailed instructions since they did not have experience with the multivariate time series data. The experimenter calibrated the eye tracker⁴ and then the participant completed two blocks of trials; the first block consisted of 10 trials and the second block consisted of 5 trials. Each trial consisted of four images displayed on the screen that contained multivariate time series data from a single device test. The participant was asked to classify the images as anomalous or normal. If the participant indicated that the image was anomalous, another screen was displayed which asked the participant to indicate the type of anomaly. Eye tracking data⁵ and response times were recorded while the subject inspected the time series stimuli.

Verbal Walkthrough. Finally, the experts and practitioners provided a Verbal walkthrough of the 15 trials. The experimenter opened a PowerPoint presentation that contained the 15 trials and asked the participant to explain their thought processes as they examined the time series data, how he/she reached their decision and what aspects of the images “popped out” or caught his/her eye for each trial. A second experimenter took notes while this discussion was taking place. The novices did not perform this task since they made decisions based on the detailed instructions that were given to them.

5 Analysis and Results

Subject response times were recorded by custom software written in Java and represent the amount of time a subject spent inspecting the time series data in each trial. Subject responses for both the anomaly/normal decision and anomaly type were also recorded by this software. Time series analysis response times trended inversely with level of experience (Fig. 3a). Experts performed best at identifying and categorizing anomalous device performance. Practitioners and novices identified anomalous device performance equally well, however practitioners performed better at anomaly categorization than did novices. Both practitioners and novices performed worse than experts at anomaly categorization (Fig. 3b).

Eye tracking fixation points and durations were calculated in the EyeWorks Analyze software by setting the pixel threshold to 45 pixels and the fixation minimum duration to 0.075 s. Given the display resolution, display size, and typical subject viewing distance (64 cm), these settings define the maximum angular velocity for fixations to be 14.4 °/s, which is consistent with previously published values [9]. An example of the screen layout and a fixation pattern are shown in Fig. 4.

⁴ FaceLAB 5 Standard System with two miniature digital cameras and one infrared illumination pod.

⁵ Eye tracking data were collected and analyzed using EyeWorks software, Eye Tracking Inc., Solana Beach, CA.

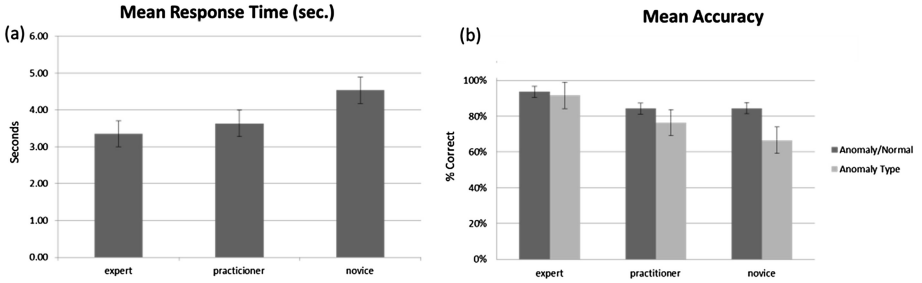


Fig. 3. Response time (a) and accuracy (b) by subject experience level for the multivariate time series task.

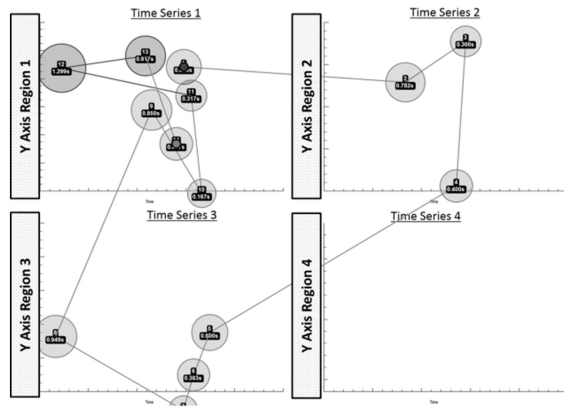


Fig. 4. Example screen layout and fixation pattern for multivariate time series task. The shaded regions, indicating the areas containing y-axis magnitude values, were used to calculate the number of times subjects fixated on the magnitude information for each axis.

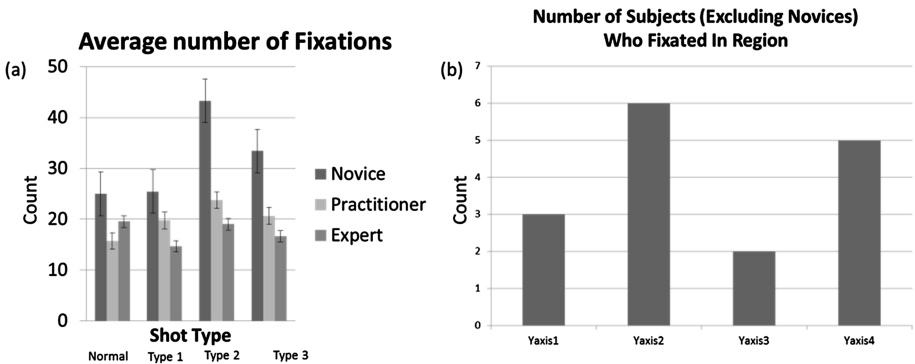


Fig. 5. Average number of fixations per trial by subject level of experience (a), and number of subjects who fixated in each Y-axis magnitude region (b).

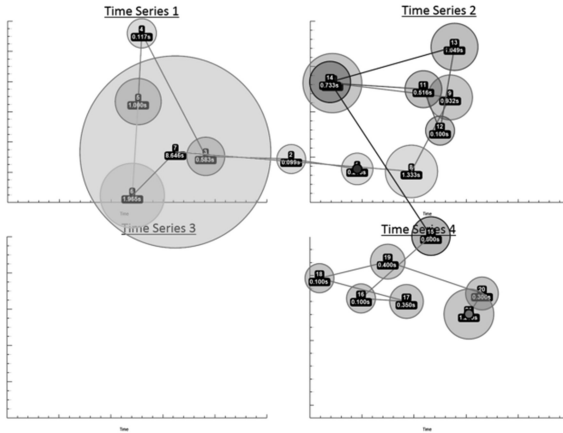


Fig. 6. Fixation pattern showing early termination of visual inspection. Note that no fixations occur in the region of Time Series 3.

In addition to the important features of the time series data, analysis of the eye tracking data revealed information about the efficiency of expertise in visual information acquisition to support abductive reasoning. As shown in Fig. 5a, experts used fewer fixations and were more consistent in the number of fixations needed to correctly identify anomalous tests and anomaly types. Fixation patterns also provide insight into the actual inspection process and types of information used to detect and diagnose abnormal device performance. These patterns were compared with analyses of the task-specific survey and verbal walkthrough to identify bias due to unconscious omission. During the verbal walkthrough, the engineers described a detailed set of time series shapes and thresholds they used to make decisions. They also provided detailed descriptions of how they reasoned about relationships across the different time series. In particular, they described their process as always inspecting each time series type and always checking certain quantitative thresholds for time series types 1, 2 and 4. While the fixation patterns corroborate the use of information from each time series type, they also showed that the engineers did not consistently fixate on y-axis magnitude values for these time series (Fig. 5b). Of the seven expert and practitioner engineers, only three fixated in the Y-axis-1 region, only six fixated on the Y-axis-2 region and only five fixated on the Y-axis-4 region. We hypothesize that the engineers have developed a pattern recognition heuristic that makes them more efficient at the analysis work, but that this heuristic is not easily verbalized. Finally, the fixation patterns showed examples of early completion of visual inspection, possibly due to satisfaction of search. As shown in Fig. 6, one subject completed a thorough inspection of three of the four time series types, but did not have any fixations in the region of time series 3. This is inconsistent with statements from this subject in the verbal walkthrough that each time series is inspected prior to making a decision. We hypothesize that this is an example of early termination of visual inspection because the subject had reached an early conclusion.

6 Summary and Conclusion

We have described a methodology for knowledge elicitation that is robust to many sources of bias. Robustness is achieved through incorporation of one or more physiological sensors to provide cross referencing information for more traditional knowledge elicitation instruments. The example application shows that eye tracking is effective at highlighting actual-ideal discrepancies that would not have been discovered by following a traditional verbal walkthrough protocol. Future work applying this methodology to additional work domains and tasks would provide the experience necessary to develop detailed guidelines for selecting physiological sensors and metrics most appropriate for a given type of task or knowledge elicitation goal.

Acknowledgements. We wish to acknowledge James D. Morrow of Sandia National Laboratories, Albuquerque New Mexico for creating the software used in our study to display the time series stimuli and record subject response times. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. Booker, J.M., Meyer, M.A.: Eliciting and Analyzing Expert Judgment: A Practical Guide. ASA, pp. 3–16. SIAM, Philadelphia (2001)
2. Thagard, P., Shelley, C.: Abductive reasoning: logic, visual thinking, and coherence. In: DallaChiara, M.L., et al. (eds.) *Logic and Scientific Methods: Volume One of the Tenth International Congress of Logic, Methodology and Philosophy of Science*, Florence, August 1995, vol. 259, pp. 413–427. Springer, Netherlands (1997)
3. Console, L., Torasso, P.: On the co-operation between abductive and temporal reasoning in medical diagnosis. *Artif. Intell. Med.* **3**(6), 291–311 (1991)
4. Kovacs, G., Spens, K.M.: Abductive reasoning in logistics research. *Int. J. Phys. Distrib. Logistics Manage.* **35**(2), 132–144 (2005)
5. Gelsema, E.S.: Abductive reasoning in bayesian belief networks using a genetic algorithm. *Pattern Recogn. Lett.* **16**(8), 865–871 (1995)
6. Brookings, J.B., Wilson, G.F., Swain, C.R.: Psychophysiological responses to changes in workload during simulated air traffic control. *Biol. Psychol.* **42**(3), 361–377 (1996)
7. Marshall, S.P.: Identifying cognitive state from eye metrics. *Aviat. Space Environ. Med.* **78**(5), B165–B175 (2007)
8. Matzen, L.E., Haass, M.J., McNamara, L.A.: Using eye tracking to assess cognitive biases: a position paper. In: *DECISIVE Workshop, IEEE Vis 2014* (2014)
9. Erkelens, C.J., Sloot, O.B.: Initial directions and landing positions of binocular saccades. *Vision. Res.* **35**(23–24), 3297–3303 (1995)