DTU Library

Towards Diagram Understanding: A Pilot Study Measuring Cognitive Workload Through Eye-Tracking

Maier, Anja; Baltsen, Nick; Christoffersen, Henrik; Störrle, Harald

Published in:
Proceedings of International Conference on Human Behaviour in Design 2014

Publication date:
2014

Document Version
Early version, also known as pre-print

Link back to DTU Orbit

Citation (APA):
Maier, A., Baltsen, N., Christoffersen, H., & Störrle, H. (2014). Towards Diagram Understanding: A Pilot Study Measuring Cognitive Workload Through Eye-Tracking. In *Proceedings of International Conference on Human Behaviour in Design 2014* Article HBiD2014/114

TOWARDS DIAGRAM UNDERSTANDING: A PILOT STUDY MEASURING COGNITIVE WORKLOAD THROUGH EYE-TRACKING

Anja M Maier[1], Nick Baltsen[1], Henrik Christoffersen [1], and Harald Störrle[2]
[1]Deptartment of Management Engineering, Technical University of Denmark
[2]Dept. of Applied Mathematics and Computer Science, Technical University of Denmark

ABSTRACT
We investigate model understanding, in particular, how the quality of the UML diagram layout impacts cognitive load. We hypothesize that this will have a significant impact on the structure and effectiveness of engineers' communication. In previous work, we have studied task performance measurements and subjective assessments; here, we also investigate behavioral indicators such as fixation and pupillary dilation. We use such indicators to explore diagram understanding- and reading strategies and how such strategies are impacted, e.g. by diagram type and expertise level. In the pilot eye-tracking experiment run so far, we have only examined a small number of participants (n=4), so our results are preliminary in nature and do not afford far reaching conclusions. They do, however, corroborate findings from earlier experiments, for example, showing that layout quality indeed matters and improves understanding. Our results also give rise to a number of new hypotheses about diagram understanding strategies that we are investigating in an ongoing data acquisition campaign.

Keywords: Design cognition, Design communication, Diagram understanding, Eye-tracking, Layout quality, Model comprehension, Model-based systems engineering, UML, Visual attention

1 INTRODUCTION

Engineering systems are becoming increasingly reliant on embedded software and thus, Software Engineering (SE) is becoming an integral part of other forms of engineering. Simultaneously, engineering design is becoming more and more model-centric [1], a practice that has been in widespread use in software development under the name Model Based Software Engineering (MBSE). There, the most widely used notation is the Unified Modeling Language (UML [2]), which can be considered the *lingua franca of software engineering* [3]. Variants of UML such as Systems Modeling Language (SysML) address the modeling challenges of a wide range of cyber-physical systems [4]. In summary, models expressed in UML and other UML-like notations are widely used in various kinds of engineering.

Models are used for a large variety of purposes, such as quality analysis, understanding, maintenance, and as a means to communicate user requirements and intended functionality or full capability of the system [5]. It is therefore fair to say that conceptual diagrams of a model representing an original, such as an idea, a process, or software, are today an important part of complex engineering systems. Diagram understanding thus becomes essential for efficient and effective systems design. We need to understand what determines diagram understanding, in order to gain insights into what affects communication between stakeholders [6].

One of the aspects that have been studied in the past is layout quality of diagrams (e.g. [7]). A good layout is one that minimizes line bends and crossings, avoids obscuring, uses visual variables sparingly and uniformly, joins similar elements, provides symmetric arrangements where applicable, and respects flow. Various empirical studies have presented evidence to confirm that well-laid out diagrams incur less cognitive load on behalf of the modelers trying to understand diagrams. Mostly, these studies have used task performance measurements, such as score, time, and subjective assessments for perceived difficulty and effort. In the work reported here, we strive to

1) replicate previous experiments using physiological measurements using eye-tracking instead of subjective assessments, and

2) explore the mental processes involved in understanding good and bad layouts, respectively, so that we arrive at a set of hypotheses that explain observations we and others before us have made.

In the remainder of this paper, we present a brief literature review on the impact of layout quality and measuring visual attention and cognitive workload through eye-tracking (Section 2). It is followed by a description of the experiment setup in Section 3, preliminary results in Section 4, and a discussion leading to new hypotheses that will have to be studied in future work. The paper is concluded in Section 5.

2 LITERATURE BACKGROUND

Diagram Layout is concerned with the way in which parts are arranged and laid out. In the context of a UML diagram, layout refers to the design and arrangement of UML elements in the diagrams. Practical experience with UML diagrams suggests that layout has a direct impact on readability. UML diagrams with a good layout will make it easier to understand the content, and therefore make it easier to work with.

Eye-tracking allows the detection and recording of a viewer's eye movements [8], indicating visual attention (where, when and for how long a subject looks at a specific spot), eye blinks and pupil dilation at a given time [9]. Analysis of eye movements while doing a task can contribute to the understanding of cognitive processes. For instance, [10] argue that the relation between eye behavior and cognition can be used to study cognitive processing during reading [11], visual search [12] and problem solving [13].

In engineering design research, pilot experiments have been conducted using eye-tracking to investigate how (mechanical) engineers go about understanding the functionality of the technical system being represented by a technical drawing [14]. Eye-tracking appears to be a suitable method to elicit insight into strategies employed by engineers reading the diagram. In software engineering, earlier research using eye-tracking suggests that inter-personal variation of diagram navigation depends on UML expertise level and other capabilities [15]. Recent research results have shown that layout quality of UML diagrams strongly influences comprehension of the underlying models [7].

Visual attention of the eyes on particular locations triggers mental processes to solve a given task [16]. Visual attention is studied through gaze fixation count, duration, and average fixation rate [17]. Another indicator for cognitive workload are changes in pupil dilation ([18],[19]). Pupil dilation (PD) is strongly related to activity in the locus caeruleus-nore pinephrine system of the brain, which is activated by stress, memory retrieval and memory consolidation [20], [21]. PD changes are usually very small (<0.5mm) [22].

3 PILOT STUDY DESIGN

Objective: Previous studies have shown that diagram layout has an effect on comprehension and that layout quality benefits understanding of what is modelled. Cognitive workload has been used as a measure for comprehension, yet, results are so far subjective. The objective of this pilot study was to measure cognitive workload using pupil dilation and fixations through eye-tracking.

Measurement equipment and measures: As eye-tracking hardware, the SMI RED 4 (FireWire) was used, together with the software SMI iViewX v2.8.26, SMI Experiment Center v3.4.148, and SMI BeGaze v3.4.148. With the AOI Editor, areas of interest (AOI) were created for the presented stimuli (the UML diagrams). Pupil diameter and fixation measures were used. Pupillometry has been suggested as the most reliable indicator of cognitive workload [23][24].

Samples: A total of nine UML diagrams were selected from a previous study [7]. The diagrams differed with respect to type (activity diagrams (AD), class diagrams (CD), and use case diagrams (UC)), layout size in terms of number of elements (small, large) and layout quality (good, bad). The experiment comprised five participants, three senior scientists with high levels of UML expertise, and two graduate students with a background in software engineering and some familiarity with UML. Due to calibration issues, only the results of four participants are included in this paper.

Experiment protocol: Following an instruction on the experiment process, calibration of the eye-tracking, and introduction on how to use the equipment and how to complete the experiment tasks, the experiment proceeded as follows: Participants were presented a sequence of nine diagrams on screen, ten comprehension task questions for each diagram and an evaluation after each diagram indicating difficulty to comprehend the diagrams and effort required to solve the task. Participants were given a time limit of two minutes to get an overview of the diagram and one minute to solve each comprehension task. The process of getting an overview of a diagram, solving the ten comprehension tasks and accessing the difficulty and effort was repeated for each of the nine diagrams in the pilot study. Demographic questions completed the experiment. Each experiment lasted ca. 30 minutes.

4 OBSERVATIONS AND INTERPRETATIONS

Table 1 shows the measurements for various dependent variables (columns) across the stimuli (rows). We show only averages here. The first set of columns shows the diagram number and information about diagram type, layout quality and size. Since our first objective was to replicate existing results, we have sampled the stimuli from earlier experiments to cover all aspects of layout that [Störrle] has examined. The next four columns show the eye-tracking results (pupil diameter, blink frequency, blink duration, fixation duration), where large pupil diameters indicate high cognitive workload [18], [19], [23]–[26], high blink frequency [19], [27] and high blink duration [28], indicate high cognitive workload. Similarly, high fixation duration indicates high cognitive workload. The next two columns show the same task performance indicators [26] as measured (time taken and accuracy score) to allow a direct comparison. The last two columns show the subjective retrospective assessments on difficulty and effort on a scale from 1-5, with 1 being low in difficulty and effort respectively.

4.1 Variable differences

The rows are sorted by increasing number of measurements indicating cognitive load above and performance below average; those values are highlighted by inverting text color. That is to say, the first row shows the measurements for the stimulus where participants exhibited the worst performance and the highest cognitive load. Informed by earlier results and our working hypotheses, we would expect increasingly fewer bold face values towards the bottom of the table. Clearly, that is the case over all. We interpret this as, more or less, confirming earlier results: large diagram size and low diagram quality correlate with low performance and/or high cognitive load.

Table 1: Classifications of Stimuli and measurements taken in the pilot study (n=4)

No.	**Layout quality**	**Diagram Type**	**Diagr. Size**	**Pupil Diam.** *[mm]*	**Blink Freq.** *[#/s]*	**Blink Dur.** *[ms]*	**Fixation Dur.** *[ms]*	**Time** *[s]*	**Accuracy** *[%]*	**Difficulty** *[1-5]*	**Effort** *[1-5]*
Bad values are:				*large*	*high*	*low*	*high*	*high*	*low*	*high*	*high*
4	*Low*	*Activity*	*Small*	**2,69**	0,07	**127,87**	**289,12**	**16,67**	**82%**	**3,60**	**3,80**
24	*Low*	*Activity*	*Large*	**2,68**	**0,10**	175,17	**296,14**	**16,18**	**86%**	**3,20**	**3,40**
33	*High*	*Use Case*	*Large*	**2,70**	0,08	**124,67**	**298,77**	**13,37**	100%	**3,00**	2,80
10	*Low*	*Class*	*Small*	2,66	**0,09**	**115,92**	276,26	11,31	**86%**	**3,20**	**3,40**
13	*High*	*Use Case*	*Small*	2,67	**0,10**	**121,67**	**288,91**	9,34	96%	2,00	1,80
1	*High*	*Activity*	*Small*	2,63	**0,09**	158,07	**287,45**	**12,88**	**82%**	2,60	**3,00**
29	*High*	*Class*	*Large*	2,64	0,06	225,77	**286,91**	11,98	88%	**3,00**	**3,40**
7	*High*	*Class*	*Small*	**2,68**	0,07	209,18	279,53	10,34	78%	2,20	1,80
18	*Low*	*Use Case*	*Small*	2,67	0,06	**119,00**	257,12	7,64	92%	2,60	2,60
Average				**2,67**	**0,08**	**156,04**	**285,81**	**12,19**	**88%**	**2,82**	**2,89**

Interestingly, the corroboration is clearer for the objective performance and subjective assessments that we had studied in earlier experiments, and not quite as clear for the cognitive load indicators. Also, there is a notable divergence between the various indicators. For instance, the two objective performance measures diverge only for two diagrams, the two subjective assessment indicators diverge on three diagrams, and all pairs of these variables diverge on 2.5 diagrams on average. In contrast, any pair of physiological measures diverges in four to six diagrams (4.8 on average). So, the results for the *observable* behavior (time, errors) and the *conscious* account of difficulty and effort are much more consistent with each other than the results of the various *physiological* measures. Since we are only comparing whether a measurement is above/below the respective average, measurement inaccuracies do not provide no likely explanation of this observation.

We believe it is more likely that the physiological measures do indeed indicate one or more *subconscious* processes are involved to varying degrees. Thus we define our first new hypothesis **(H1): there are several different mental processes involved in diagram understanding which are reflected in consistently diverging measures of physiological variables for different kinds and problems of diagram layouts.** We also hypothesise that **(H2) there are different strategies for diagram understanding that will lead to different profiles in the various physiological measures.**

4.2 Layout style

Based on our intuition and introspection of our own actions in creating and understanding UML diagram layouts, we identify certain common layout patterns. For instance, sequential layouts following a natural flow direction (top-bottom, left-right, clockwise) is often used in activity and state diagrams. Similarly, a hub-and-spoke pattern is often seen in class and use case diagrams focusing on a given element that is placed centrally in a layout. Initial inspection of the saccadic movements and fixation sequences (not shown in this paper) indicate that the usage of such layout patterns is reflected by analogous eye movements. We thus hypothesize that **(H3) if a diagram follows a common layout pattern, the pattern is also discovered in modelers' eye-movements and fixation patterns.** We also assume that (h4) if the layout pattern has a natural start or end point (e.g., an InitialNode place at the top of a diagram, or a central class placed in the middle of a class diagram), then modelers will start or end there, and are associated with higher cognitive load.

The initial data also suggests that there is no obvious pattern with regards to diagram type, i.e., all diagram types contain "easy" and "hard" parts; no single diagram type is "hard" as such, judging by its diagram elements. For instance line crossings are easily identified as difficult in all types of diagrams studied (2^{nd} largest pupil diameter and 3^{rd} largest fixation duration), whereas Comment boxes have less influence, independent of diagram type. Thus, we hypothesize that **(H4) the diagram type has no influence on modeler performance, but layout style does.**

4.3 Grapheme Differences

We have also measured pupil diameter and fixation duration for different element kinds. This gives rise to a ranking of the elements, indicating their relative weight according to the two variables pupil diameter and fixation duration. The results are shown in Table 2 below. Clearly, the rankings obtained for the two variables are quite different, suggesting they measure different phenomena.

We may categorize the elements according to whether they derive their meaning primarily from the label attached to them, or purely from their shape and their connections in the overall graph. For instance, UseCases[1] and Actions are represented by ellipses and roundtangles, respectively, whose dominant feature is the label placed inside the shape. The difference between any two Actions or UseCases, respectively, arises solely from this label. Conversely, MergeNodes and InitialNodes usually have no label, and their meaning primarily derives from their connections to other nodes. This type of node can further be subdivided into elements that imply a certain direction of the layout and reading flow like InitialNode and the Includes-relationship and those that represent a one-to-many split such as MergeNodes and ForkNodes.

It appears that the elements in the latter group are associated to relatively high fixation durations, but not to large pupil diameters, whereas the elements in the former group show the exact opposite. This might indicate that different processes are involved in processing (these) different kinds of elements. Possibly, the fixation duration is increased for graphemes that require a time-consuming process, e.g., one that has to consider several alternatives, as is the case for a set of alternating branches: each of them must be evaluated in some way though the evaluation as such does not imply an increase in cognitive load. We thus hypothesize that **(H5) branches will consistently be fixated on longer than other nodes.** On the other hand, there are elements that seem to be associated with increased cognitive load but not prolonged fixation, indicating that they are processed as a single unit, which reinforces hypothesis 3 from above.

[1] We use the canonical CamlCaps notation for UML meta-classes, following the UML standard.

Table 2: Diagram elements based on avg. values for pupil diameter and fixation duration.

Diagram Element	Diagram Type	Relevance for	Pupil diameter		Fixation duration	
			[mm]	Rank	[ms]	Rank
Action	Activity	-	2.69	7	339.85	5
Actor	Use Case	-	2.65	17	238.22	19
Class	Class	-	2.67	12	283.83	15
DatastoreNode	Activity	Split	2.72	5	286.29	14
DecisionNode	Activity	Split	2.7	6	403.7	1
FinalNode	Activity	Flow	2.86	1	311.38	7
ForksNode	Activity	Split	2.63	19	271.38	18
Forks_out	Activity	Split	2.6	20	311.15	8
Headings	Any	-	2.65	16	273.57	16
Includes	Use Case	Flow	2.64	18	272.06	17
InitialNode	Activity	Flow	2.65	15	343.75	4
Linecrossings	Any	Flow	2.73	3	380.52	2
MergeNode	Activity	Split	2.68	10	345.88	3
Multiplicity	Class	-	2.67	11	329.38	6
Comment	Any	-	2.65	14	290.57	11
DataFlowNode	Activity	Flow	2.65	13	288.09	12
Subject	Use Case	-	2.72	4	293.6	9
Swimlane	Activity	Flow	2.68	9	227.31	20
System	Use Case	-	2.82	2	286.8	13
UseCase	Use Case	-	2.68	8	291.62	10

5 CONTRIBUTIONS, LIMITATIONS, AND CONCLUSIONS

5.1 Summary

We have pursued two main goals with the research reported here, replicating earlier experiments to corroborate their results and generating five new hypotheses. We have replicated previous experiment with new methods, participants, and experimenters, and, to a certain degree, corroborated our earlier findings. Of course, the small sample size of this pilot study does not yet support any far reaching conclusions; the results of the ongoing larger experiment might provide that. Finally, we wanted to use eye-tracking data to inform and explore potential explanations of mental strategies for diagram understanding and the factors influencing such strategies.

5.2 Scope and Limitations

Probably the most serious limitation of our work is the small sample sizes in terms of participants and the stimuli used in the study. While the sample size is in line with studies reported in much of the literature on using eye-tracking in UML diagram understanding, we believe the sample is too small to support far-reaching conclusions. Thus, we consider this a pilot study that paved the way for a larger, currently ongoing study. It helped us to validate the experimental procedure and inform the creation of new hypotheses.

5.3 Future Work and Implications

So far, we have only studied the impact of layout quality on understanding by individuals. We think, however, that it is even more interesting to see whether and how layout quality affects communication: how do engineers 'read' and 'navigate' technical diagrams, such as UML diagrams, in their practical work? Using eye-tracking "in the field", however, poses a whole new set of challenges. Greater insight into diagram understanding strategies will allow us to create more effective diagrammatic representations, improving engineering design in general and engineering design communication in particular.

REFERENCES

[1] A. M. Maier, D. C. Wynn, T. J. Howard, and M. M. Andreasen, “Perceiving design as modelling: A cybernetic systems perspective,” in in *An Anthology of Theories and Models of Design*, L. T. M. Chakrabarti, Amaresh, Blessing, Ed. Springer, 2014, pp. 133–148.

[2] O. M. G. (OMG), “Unified Modeling Language™ (UML®) Resource Page.” [Online]. Available: http://www.uml.org/.

[3] B. Selic, “The pragmatics of model-driven development. Software,” *IEEE Softw.*, vol. 20, no. 5, pp. 19–25, 2003.

[4] O. M. G. (OMG), “OMG Systems Modeling Language (OMG SysML™),” 2012.

[5] A. Maier, D. C. Wynn, M. M. Andreasen, and P. J. Clarkson, “A cybernetic perspective on methods and process models in collaborative designing,” in *Proceedings of the 12th International Design Conference DESIGN 2012*, 2012, pp. 233–240.

[6] A. M. Maier, M. Kreimeyer, C. Hepperle, C. M. Eckert, U. Lindemann, and J. P. Clarkson, “Exploration of correlations between factors influencing communication in complex product development,” *Concurr. Eng. Res. Appl.*, vol. 16, no. 1, pp. 37–59, 2008.

[7] H. Störrle, “On the Impact of Layout Quality to Understanding UML Diagrams: Diagram Type and Expertise,” in *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2012, pp. 49–56.

[8] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. 2007.

[9] L. Maughan, S. Gutnikov, and R. Stevens, “Like more, look more. Look more, like more: The evidence from eye-tracking,” *J. Brand Manag.*, vol. 14, no. 4, pp. 335–342, 2007.

[10] E. De Kock, J. Van Biljon, and M. Pretorius, “Usability evaluation methods : Mind the gaps,” in *Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, 2009, pp. 122–131.

[11] A. Aula, P. Majaranta, and K.-J. Räihä, “Eye-tracking reveals the personal styles for search result evaluation,” in in *Human-Computer Interaction*, 2005, pp. 1058–1061.

[12] R. Bednarik and M. Tukiainen, “An eye-tracking methodology for characterizing program comprehension processes,” in *Proceedings of the 2006 symposium on Eye tracking research & applications - ETRA ’06*, 2006, pp. 125–132.

[13] P. Kasarskis, J. Stehwien, J. Hickox, and A. Aretz, “Comparison of Expert and Novice Scan Behaviors During VFR Flight,” in *Proceedings of the 11th International Symposium on Aviation Psychology*, 2001.

[14] M. M. Matthiesen S, Meboldt M, Ruckpaul A, “Eye tracking, a method for engineering design research on engineers’ behavior while analyzing technical systems,” in *Proceedings of the 19th International Conference on Engineering Design*, 2013, pp. 277–286.

[15] S. Yusuf, H. Kagdi, and J. I. Maletic, “Assessing the Comprehension of UML Class Diagrams via Eye Tracking,” in *15th IEEE International Conference on Program Comprehension (ICPC ’07)*, 2007, pp. 113–122.

[16] M. A. Just and P. A. Carpenter, “A theory of reading: from eye fixations to comprehension,” *Psychol. Rev.*, vol. 87, no. 4, pp. 329–354, Jul. 1980.

[17] B. Sharif, “Empirical assessment of UML class diagram layouts based on architectural importance,” in *2011 27th IEEE International Conference on Software Maintenance (ICSM)*, 2011, pp. 544–549.

[18] D. Kahneman, *Attention and Effort*. Englewood Cliffs, New Jersey: Prentice Hall Inc, 1973.

[19] G. F. Wilson, “Operator functional state assessment,” North Atlantic Treaty Organisation, Report No. AC/323(HFM-104)TP/48, 2004.

[20] G. Aston-Jones and J. D. Cohen, “An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance,” *Annu. Rev. Neurosci.*, vol. 28, pp. 403–450, 2005.

[21] B. Laeng, S. Sirois, and G. Gredeback, “Pupillometry: A Window to the Preconscious?,” *Perspect. Psychol. Sci.*, vol. 7, no. 1, pp. 18–27, Jan. 2012.

[22] A. Bojko, *Eye-Tracking the User Experience: A practical guide to research*. Brooklyn, New York: Rosenfeld Media, 2013.

[23] J. Beatty and B. Lucero-Wagoner, “The Pupillary System,” in *Handbook of Psychophysiology*, 2000.

[24] N. Meshkati, P. A. Hancock, M. Rahimi, and S. M. Dawes, “Techniques in mental workload assessment,” in in *Evaluation of human work: A practical ergonomics methodology*, 1995, pp. 749–782.

[25] J. Beatty, “Task-Evoked Pupillary Responses, Processing Load and the Structure og Processing Resources,” *Psychol. Bull.*, vol. 91, no. 2, pp. 276–292, 1982.

[26] S. P. Marshall, “The Index of Cognitive Activity : Measuring Cognitive Workload,” in *Proceedings of the 2002 IEEE 7th conference on Human factors and power plants*, 2002, pp. 5–9.

[27] M. C. Castor, *GARTEUR Handbook of mental workload measurement*. Report No. Garteur TP145, 2003.

[28] B. Cain, “A Review of the Mental Workload Literature,” North Atlantic Treaty Organisation, Report No. RTO-TR-HFM-121-Part-II, 2007.