Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2012 Proceedings | Proceedings

Eye Movements, Perceptions, and Performance

Soussan Djamasbi
Research Laboratory, Worcester Polytechnic Institute, Worcester, MA, United States., djamasbi@wpi.edu

Ami Samani
Research Laboratory, Worcester Polytechnic Institute, Worcester, MA, United States., amisamani@wpi.edu

Dhiren Mehta
Research Laboratory, Worcester Polytechnic Institute, Worcester, MA, United States., dhiren.mehta@wpi.edu

Eye Movements, Perceptions, and Performance

Soussan Djamasbi
UXDM Research Laboratory
Worcester Polytechnic Institute
djamasbi@wpi.edu

Dhiren Mehta
UXDM Research Laboratory
Worcester Polytechnic Institute
dhiren.mehta@wpi.edu

Ami Samani
UXDM Research Laboratory
Worcester Polytechnic Institute
amisamani@wpi.edu

ABSTRACT

Due to the ever growing amount of information, individuals often face high cognitive loads when making decisions. Thus, understanding user reactions to high cognitive loads can help to improve a user's ability to make good quality decisions under high cognitive load. Prior decision making and user experience research suggests that eye tracking may provide a more complete picture of user reactions under high cognitive loads. Thus, through an exploratory study we investigate the relationship between fixation, perceptions of cognitive load and performance. Our analysis shows that fixation can predict both perception of the load as well as performance of a cognitively demanding online game.

Keywords (Required)

Eye tracking, Fixation, Cognitive Load, Online Games, Performance, Human Computer Interaction (HCI)

INTRODUCTION

Today's competitive business environment requires decision makers to process large amount of information in short periods of time. This in turn forces decision makers to work under high cognitive loads. Consequently, understanding users' behavior under high cognitive loads has become a major focus in information systems research (Djamasbi, Tulu, Loiacono, and Whitefleet-Smith, 2008; Payne, Bettman, and Johnson, 1988; Schroder, Driver, and Streufert, 1967; Svenson, 1993). For example, studies show that extremely demanding decision environments have a detrimental effect on performance (Schroder et al. 1967). Such investigations provide opportunity to look into factors that can predict behavior. Despite providing a continuous stream of data, eye tracking is rarely used in IS behavioral research. In particular, little work has been done to test whether objective measures of eye movement can predict performance and/or behavior. To address this gab, this study examines whether a user's fixation can predict his or her performance as well as subjective evaluations. Grounded in previous eye tracking studies (Djamasbi, Siegel, Skorinko, and Tullis, 2011) as well as decision making studies (Djamasbi et al., 2008; Payne et al., 1988; Schroder et al., 1967) we argue that fixations are likely to be a reliable predictor of behavior. To test this possibility, we conducted an exploratory laboratory study using eye tracking equipment.

BACKGROUND

Our theoretical argument is grounded in prior research that examines the relationship between cognitive load and performance. The relationship between cognitive load and performance has been studied in many different studies. For example some studies have looked at the impact of cognitive load on individual's response time (Luce, 1986), some other examined the influence of load on flexibility in adapting to a decision environment (Payne et al., 1988) and some investigated the impact of cognitive load on information processing capability (Schroder et al., 1967). These studies show that cognitive load can have a significant impact on an individual's performance. A known way to assess the impact of cognitive load is by using subjective measures (Rubio, Díaz, Martín, and Puente, 2004).

Studies suggest that cognitive load can also impact fixation. A prior eye tracking study (Ikehara and Crosby, 2005) has shown that cognitive load can affect eye movement. Unfortunately this study does not clarify how eye movement is defined due to the limitation of measuring dynamic content. A more recent eye tracking study uses the number of fixations that last longer than 300 ms as a measure of cognitive effort (Djamasbi et al., 2011). While this study does not use cognitive effort as a direct measure of load, its results suggest that fixations can serve as a suitable measure of cognitive load. This is because according to this previous study, fixation on the pages that were more cognitively demanding was indeed significantly more than fixation on the pages that were less cognitively demanding. In addition to the number of fixations, there is also evidence that cognitive load is likely to impact the length of fixation. For example, a previous study using fixation data to find the correlation between a user's gaze and complexity and the difficulty of the task shows a positive correlation between fixation duration and cognitive load (Rayner, 1998).

The above discussed studies show that cognitive load can affect performance and fixation. However, these studies have not examined whether there is a link between fixation and performance. Nor have they examined the impact of load in such relationship. Additionally, the perception of cognitive load is often captured via self-report measures such as the NASA Task load Index (TLX) survey. Because cognitive load can affect fixation it is likely that fixation can predict the perception of task load. However, little work has been done to examine such relationship between fixation and self-report measures of cognitive load. To explore these possibilities we conducted an eye racking laboratory experiment.

METHOD

Participants and Design

Thirteen undergrad students from an MIS course in a major university were recruited to participate in a within-subjects controlled experiment. All of the participants were experts in using computers but none was familiar with the task (the online game) used in the experiment. The participants were assigned to perform the same task twice at different cognitive loads. The order of level of cognitive load assigned was random meaning alternate students were asked to perform the task with lower cognitive load first followed by a higher load cognitive task. This experimental setting allowed for uniformity of cognitive load, eye tracking and performance measurement across all participants. None of the participants had prior knowledge of the task. We ensured that we did not disclose the specifics of the task to the participants prior to the experiment. None of the participants acknowledged having performed the task prior to this one. Consistent instructions and approximately 2 minutes of training were given to the participants prior to the actual experiment. The training familiarized the participants with the task.

Procedure

The experiment was conducted in a user experience laboratory, which was designed to mimic a typical work environment. Attached to the monitor on the desk, however, was an eye tracker. Data was collected for each participant individually. Each data collection session lasted approximately 30 minutes.

At the beginning of each session a questionnaire was provided to participants to fill out their demographic information, familiarity with computers. Next, the experimenter calibrated the eye tracker for the participant. This is a brief procedure during which the participant's gaze is mapped to several points on the screen. After calibration, participants received instructions about the task and were asked to complete one practice trial which took about 2 minutes. Next, the participant was asked to play an online game for five minutes at two levels of difficulty: intermediate and expert level. After each game the participant was asked to fill the NASA Task load Index (TLX) survey, which captures one's perception of the task's cognitive load.

Task

The task consisted of an online logic game named 'Netwalk' (http://www.logicgamesonline.com/netwalk/). The Netwalk game has several scrambled pipes, computers, and a central server. The participant was asked to behave as a network administrator and connect the pipes to join the computers to the central server such that there are no loose ends in the network. The pipes could be rotated clockwise using left click and right click was to be used to rotate a tile clockwise.

The game used in the study had three levels of difficulties; the different difficulties corresponded to different sizes of the network board. The Beginner level corresponded to a 5x5 layout, medium to a 7x7 layout and an Expert to a 9x9. Additionally, the expert level had an added difficulty. The connections in the expert level could wrap from left to right and top to bottom. The beginner level was used as the practice trial. The intermediate and expert levels were used as the actual trials in the task. This arrangement allowed us to provide the participants with a suitable practice trial at a lower cognitive load so that they can become comfortable with the logistics of playing the game. The arrangement also allowed us not to expose participants to any of the actual load levels that were used in the task while practicing the game.

Measurements

Previous studies have shown the importance and relevance of tracking fixation data, eye movements as a useful measure of cognitive load (Rayner, 1998). We used the Tobii X120 eye tracking device in order to capture fixation data. The X120 eye tracker is a standalone eye tracker that can be placed in front of a monitor for tracking a user's eye movement when using the display.   A main computer was paired with the eye tracker to collect data.

In this study, we used two types of eye tracking data: fixation count and fixation length. Fixations refer to gazes that last at least for 300 milliseconds (Djamasbi, Siegel, and Tullis, 2010). Fixation count refers to the number of times a person had a fixation on the screen and fixation length refers to the duration of each fixation, i.e., steady gazes that were longer than 300 ms (Djamasbi et al., 2011).

The NASA Task load Index (TLX) was used to measure subjective load. NASA-TLX is designed to obtain a feedback on the workload from the participants after the task. We chose TLX to measure subjective cognitive load because it has been used successfully in a number of usability studies to evaluate cognitive load (Lin and Imamiya, 2006).

Another way to assess one's cognitive load is by examining the objective measure of performance (Schroder et al., 1967). The objective measure of performance in our study was calculated by counting the number of moves the participants made in order to complete the task. This number reflected the exact number of changes the participant made to change the direction of game pieces (pipes on this case) by eliminating moves that did not have any effect on the position of the pipes. This measure is one of the two performance criteria by the game (completion time and number of moves). Because almost all the users took the whole allocated time to complete the task, using the number of moves served as a suitable measure of performance in our study.

RESULTS

First we tested the manipulation to make sure that the two cognitive levels were significantly different. This was done through to paired t-tests. The first test compared the objective measure of performance, namely the number of moves to complete the task. The second test compared the differences in subjective measure of load, namely the TLX survey. As shown in Table 1, both tests were significant indicating that performance was significantly worse in the high load condition, and that subjects perceived the high load condition to be significantly more demanding.

	Intermediate load	High load	
Objective measure of load (performance)	39.85 (16.27)	67.62 (17.83)	df= 12, t-Stat= 8.42, p=0.000
Subjective measure of load (TLX)	5.32 (1.87)	6.66 (1.88)	df= 12, t-Stat= 2.97, p=0.012

Table 1: Manipulation check. Results of the paired t-test comparing objective measure of performance as well as the subjective measure of cognitive load in the two experimental conditions

We then tested to see whether fixation data could predict performance. Because fixation count and fixation length are correlated we ran two separate regressions to conduct these investigations. Thus we used the following two models:

Performance= a0+ b1*Fixation Count+b2*Cognitive Load+ b3*Fixation Count* Cognitive Load (1)

Performance= a0+ b1*Fixation Lenght+b2*Cognitive Load+ b3*Fixation Length * Cognitive Load (2)

The only significant component in the first model (Equation 1) was fixation count. This result showed that regardless of cognitive load fixation count was a good predictor of performance. The larger the number of fixations the worse was the performance. We removed the non-significant factors and ran the model again. The results showed that 34% of the variation in this regression model was explained by the fixation count.

Dependent Variable	Independent Variable	B	t-Value	P-value
Performance	Fixation Count	0.22	2.74	0.012
Overall model F = 12.23; p = 0.000; R^2 = 0.34; adjusted R^2 = 0.31				

Table 2: Regression results for the relationship between performance and fixation count (Equation 1)

The regression model investigating the relationship between fixation length and performance (Equation 2) showed no significant results. As customary in exploratory analysis, we continued running the model by removing the non-significant results one by one. This process revealed that only the interaction between the cognitive load and fixation length had a significant impact on performance. The results showed that 45% of variance in the model was explained by this interaction. This result showed that under the high load, longer fixations were indicators of worse performance; this was not the case under the intermediate cognitive load condition.

Dependent Variable	Independent Variable	B	t-Value	P-value
Performance	Fixation Length* Cognitive Load	0.11	4.45	0.01
Overall model F = 19.83; p = 0.000; R^2 = 0.45; adjusted R^2 = 0.43				

Table 3: Regression results for the relationship between performance and fixation length (Equation 2)

The above analysis examined the relationship between the objective measures of performance and fixations. Next, we examined to see if we found any relationship between the self-report measures and eye movements. We used the TLX survey to capture users' perception of cognitive load. Again we ran two separate regressions using the following models:

TLX= a0+ b1*Fixation Count+b2*Cognitive Load+ b3*Fixation Count* Cognitive Load (3)

TLX= a0+ b1*Fixation Lenght+b2*Cognitive Load+ b3*Fixation Length * Cognitive Load (4)

We did not find any main effects for fixation count and length in the above two models. However, we found significant interaction effects in both models (Equation 3 and 4). The interaction between fixation length and load had a significant impact on TLX. Under high load the number of fixations was positively related to TLX score while this was not true under the intermediate load. The same interaction effect was also found for the impact of load and fixation count on TLX. However the p-value for this interaction effect in Equation 4 was only marginally significant (p=0.052).

Dependent Variable	Independent Variable	B	t-Value	P-value
TLX	Fixation Count X Cognitive Load	0.005	2.04	0.05
Overall model F = 4.15; p = 0.05; R^2 = 0.14; adjusted R^2 = 0.11				

Table 4: Regression results for the relationship between TLX and fixation count (Equation 3)

Dependent Variable	Independent Variable	B	t-Value	P-value
TLX	Fixation Length X Cognitive Load	0.007	5.28	0.03
Overall model F = 5.28; p = 0.03; R^2 = 0.18; adjusted R^2 = 0.15				

Table 5: Regression results for the relationship between TLX and fixation length (Equation 3)

DISCUSSION

In this study, our goal was to explore the various possible methods to interpret the relationship between cognitive load, performance and fixation. We also wanted to establish whether a subjective measure of cognitive load can be predicted by objective measures like fixation count and length.

Because we were interested in high cognitive loads we chose two demanding levels of the same game for our study. The results confirmed that the two levels of the game in this study were indeed cognitively demanding. The results also verified that the cognitive load in the two different game levels were significantly different. This was not only evidenced in the self-reported measure of TLX but also from the significant differences in the objective measure of performance in the two games completed by the participants in our study.

The results also showed a significant relationship between fixation count and performance. The higher fixation count was correlated with higher number of moves or worse performance. While the results showed that fixation count had a significant main effect on performance, only an interaction effect between cognitive load and fixation length was found to be significant.

The results also showed a significant interaction between objective load and fixation influenced the participant's subjective evaluation of the task cognitive load. These results indicated that under high cognitive load, fixation (both fixation count and fixation length) was a good predictor of the users' self-reported measure of load (TLX). In other words, these results showed a significant relationship between the subjective and objective measures of cognitive load.

The results of this study have important contributions to human computer interaction (HCI) research. They provide evidence that a user's experience of an online game can be measured through their physiological measure of fixation. Fixation can predict both the perception of the cognitive load as well as performance. Because fixation provides a continuous measure of reaction, it can provide a more complete picture of a user's reaction and thus help researchers to refine their analysis of user's behavior and reactions to a system.

The results have important implication for practice because they provide insight for designers. Using fixations in usability testing is likely to help improve the design of online games because they can provide detailed reactions that otherwise not available. A recent previous study (Djamasbi et al., 2011) has shown that visual design can affect the number fixations. Because our results showed that the number of fixation was an indicator of performance, designer may be able to reduce number of fixation (and thus improve the performance) by manipulating the visual design of their game.

LIMITATIONS AND FUTURE RESEARCH

Laboratory experiments facilitate the necessary environment for controlling desired variables and thus help us to more precisely manipulate, control, and measure their effects (Staw and Barsade, 1993; Swieringa and Weick, 1982). In particular, conducting this study as a laboratory experiment allowed us to track users' eye movements and examine the effects of cognitive load on their fixation behavior. As with all laboratory experiments, however, the generalizability of our results is limited by the experimental setting and the task we used in our study. Future studies using various tasks and environments are needed to increase our confidence in the generalizability of these results. In particular, in this study, we examined the impact of cognitive load on performance on an online puzzle game. Future studies should examine whether the results observed in our study also apply to other types of games or tasks other than games.

CONCLUSION

The results of this study support prior findings that suggest including eye tracking in user experience research can help to better understand users' reactions. Our results showing a significant relationship under high cognitive load between the subjective and physiological measures of load as well as between physiological measures of fixations and performance provide further rationale and strong support for using eye tracking in the user experience research. From a practical point of view, our study provides insight for designers indicating that paying attention to fixation count may help them design more successful games.

REFERENCES

1. Djamasbi, S., Siegel, M., Skorinko, J., and Tullis, T. (2011). Online Viewing and Aesthetic Preferences of Generation Y and the Baby Boom Generation: Testing User Web Site Experience Through Eye Tracking. *International Journal of Electronic Commerce, 15*,4, 121-158.
2. Djamasbi, S., Siegel, M., and Tullis, T. (2010). Generation Y, web design, and eye tracking. *International journal of human-computer studies, 68,*5, 307-323.
3. Djamasbi, S., Tulu, B., Loiacono, E., and Whitefleet-Smith, J. A. (2008). Can a Reasonable Time Limit Improve the Effective Usage of a Computerized Decision Aid? *Communications of the Association for Information Systems, 23,* 22,393-408.
4. Ikehara, C. S., and Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. *Proceedings of the 38th annual hawaii international conference on system sciences*, January 3-6, 295
5. Lin, T., and Imamiya, A. (2006). Evaluating usability based on multimodal information: an empirical study. *Proceedings of the 8th international conference on multimodal interfaces*, November 2-4, Alberta, Canada.
6. Luce, R. D. (1986). Response times: their role in inferring elementary mental organization: Oxford University Press, New York, USA.
7. Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,*3, 534-552.
8. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*,3, 372.
9. Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology, 53*,1, 61-86.
10. Schroder, H. M., Driver, M. J., and Streufert, S. (1967). Human information processing: New York: Holt, Rinehart and Winston, 54-61.
11. Staw, B. M., and Barsade, S. G. (1993). Affect and managerial performance: A test of the sadder-but-wiser vs. happier-and-smarter hypotheses. *Administrative Science Quarterly*, 38,2, 304-331.
12. Svenson, O. (1993). Time pressure and stress in human judgment and decision making: Springer, Pletinum Press, New York, USA
13. Swieringa, R. J., and Weick, K. E. (1982). An assessment of laboratory experiments in accounting. *Journal of Accounting Research, 20*, 56-101.