# Saccadic Eye Movements and Search Task Difficulty as Basis of Modeling User Knowledge in Information Seeking

**Ying-Hsang Liu (ying-hsang.liu@abm.uu.se)[1,3]**
**Andreas Nürnberger (andreas.nuernberger@ovgu.de)[2]**
**Jenny Rettstatt (jenny.rettstatt@hsw.tu–chemnitz.de)[3]**
**Marco Ragni (marco.ragni@hsw.tu–chemnitz.de)[3]**
[1] Department of ALM, Uppsala University, Sweden
[2] Data and Knowledge Engineering Group, Otto von Guericke University Magdeburg, Germany
[3] Predictive Analytics, Chemnitz University of Technology, Germany

**Abstract**

Designing user-adaptive search systems necessitates modeling the user's knowledge state during information seeking. Gaze data offers insights into cognitive processes during task-based reading. Despite its potential, cognitive perspectives have been insufficiently explored in the representation of the user's knowledge state when designing search systems. We reanalyzed an eye-tracking dataset and constructed mixed-effects user models to identify which measurements of gaze activities (i.e., gaze metrics captured by eye trackers) are reflective of the user. Our study's findings indicate that there are statistically significant correlations between gaze metrics that measure the variability of saccadic eye movement and search performance. The accuracy of answers has been significantly influenced by the interaction between the control of saccade trajectories, measured by the standard deviation of absolute saccadic directions and the difficulty of the search task. We discuss the implications of these findings for the design of search systems adaptable to the user's state of knowledge.

**Keywords:** User modeling; Gaze data; Information seeking; Question-answering; Saccadic eye movements

## Introduction

Eye movements provide a window to the cognitive processes in decision making in which the accuracy of decision outcome is moderated by visual and task contexts (e.g., König et al., 2016; Shubi & Berzak, 2023; Spering, 2022). Since it is the human that searches for information, a search engine is a tool that can either adapt itself to the specific cognitive processes of the human or require the human to adapt to the engine. There are specific ways in which humans acquire information and how they form mental models and the impact of mental models on explaining search behavior (e.g., Gwizdka & Dillon, 2020; Thomas, Billerbeck, Craswell, & White, 2019). The human ability for information seeking and its specifics has been modeled for a long time (W.-T. Fu & Gray, 2006; W. T. Fu, 2020). Researchers have approached cognitive modeling of users by examining the reasoning processes using eye movements data (Purcell, Roberts, Handley, & Howarth, 2023), modeling reading comprehension in information seeking (Shubi & Berzak, 2023) and cognitive processes of adopting expert interaction techniques (Bailly, Khamassi, & Girard, 2023; Castner, Frankemolle, Keutel, Huettig, & Kasneci, 2022). Given that human gaze in reading comprehension is conditioned by visual and task contexts (Malmaud, Levy, & Berzak, 2020), it is important to specify which measurements of general gaze activities based on eye trackers (also known as gaze metrics) are most promising for predicting search success in the context of goal-oriented information seeking, such as task-based reading.

Gaze data has been used to infer user knowledge level by identifying eye movement patterns (Cole, Gwizdka, Liu, Belkin, & Zhang, 2013) and to predict user cognitive abilities (Conati, Lallé, Rahman, & Toker, 2020). Within this thread of research, search task difficulty has been identified as one of the key variables affecting user search behavior (e.g., Kim, 2005; J. Liu, Liu, Cole, Belkin, & Zhang, 2012). For example, informed by Information Foraging theory (Pirolli & Card, 1999), research findings have revealed that user-perceived search task difficulty is correlated with eye movements in specified areas of interest (AOI) in search interfaces and search performance (Wittek, Liu, Darányi, Gedeon, & Lim, 2016; Y.-H. Liu, Thomas, Gedeon, & Rusnachenko, 2022). Since previous studies have demonstrated the relevance of gaze metrics toward visual exploration (making a saccade to a new object) or exploitation (revisiting an object that was previously fixated) in information-seeking tasks (e.g., König et al., 2016; Spering, 2022), it is important to examine the relationships among search task difficulty, gaze metrics, and search performance holistically.

To that end, this study aims to answer the overall research question of how to model users' state of knowledge through eye-tracking data when they seek information via search systems. We formulate the following research questions based on previous research:

- How do we specify which gaze metrics (i.e., measurements of gaze activities captured by eye trackers) are reflective of user knowledge in task-based reading?
- What are the effects of gaze metrics, search task difficulty, and user self-perceived prior knowledge about search tasks on search performance in the context of information seeking?

Our research findings suggest that there are statistically significant correlations between gaze metrics of the SD (standard deviation) of absolute/relative saccadic directions and search performance measured by the answer correctness in fact-finding search tasks. Importantly, the search task



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

difficulty and its interactional effect with the SD of absolute saccadic directions have significantly contributed to the answer's correctness.

The major contributions of the study include:

- Identifying the correlations between gaze metrics and search success in information-seeking tasks that involve task-based reading;
- Constructing mixed-effects models for the effect of search task difficulty and gaze metrics on search success in fact-finding tasks.

## Related Work

Adaptive search systems are designed to guide the users to accomplish search tasks. For instance, research has explored the adaptive guidance in narrative visualizations by directing attention to salient components of the narrative visualizations (Barral, Lallé, & Conati, 2020). Research has also investigated the use of interaction data as an alternative to eye-tracking data for predicting cognitive abilities for user modeling in interactive visualizations (Conati et al., 2020). These studies show that adaptive guidance can support users with varying levels of visualization literacy (Barral et al., 2020), interaction data can predict cognitive abilities more accurately at the beginning of a task compared to eye-tracking data (Conati et al., 2020), and user behavior, including eye gaze and click data, can be used to predict user success in ontology class mapping tasks (B. Fu & Steichen, 2022). Further, analysis of gaze data has demonstrated that there were positive correlations between reading strategies and search activities. Specifically, hard reading was correlated with exploratory tasks, whereas skimming was correlated with fact-finding tasks (Schwerdt, Kotzyba, & Nurnberger, 2021). These findings suggest that adaptive search systems can be further developed based on the examination of gaze data in search activities.

Methodologically, machine learning models have been trained to predict the perceived relevance of paragraphs based on eye movements (Barz, Bhatti, & Sonntag, 2022). Research has shown that multimodal classifiers combining interaction data and eye-tracking data show promising results for predicting cognitive abilities (Conati et al., 2020), and gaze data trained by deep-learning classifiers can predict users' success in visual search tasks (Spiller et al., 2021).

Using statistical techniques, regression models were constructed to predict a user's level of knowledge based on real-time measurements of eye movement patterns during a task session (Cole et al., 2013). Interestingly, a unimodal based on the behavioral signal of left click in LR (Logistic Regression) performed the best for implicit detection of document relevance (González-Ibáñez, Esparza-Villamán, Vargas-Godoy, & Shah, 2019). Importantly, LR has provided the best results in a few classification experiments (e.g., Steichen, Conati, & Carenini, 2014; Raptis et al., 2017; González-Ibáñez et al., 2019). Advanced statistical techniques, such as mixed-effects models have been gradually applied in information retrieval research for the analysis of gaze behavior (Hofmann, Mitra, Radlinski, & Shokouhi, 2014), user characteristics (C. Liu et al., 2019) and search performance (Y.-H. Liu et al., 2023). Therefore, this study will use mixed-effects models to shed light on the gaze features that could predict the search success in question-answering tasks.

In consideration of previous research findings, this study aims to explore how eye-tracking data can represent users' knowledge during information seeking by identifying gaze metrics reflective of user knowledge in task-based reading, and examining the effects of gaze metrics, search task difficulty, and users' self-perceived familiarity with search tasks on their performance. We hypothesize that search task difficulty and gaze metrics affect user search performance.

## Methods

This study reanalyzed the user interaction data collected from a controlled user experiment in an academic research environment (Kotzyba, Gossen, Schwerdt, & Nürnberger, 2017; Schwerdt, Kotzyba, & Nurnberger, 2018; Schwerdt et al., 2021). A total of 19 subjects participated in the study (13 male & 6 female) recruited from mailing lists. They were mostly young professionals (32.4 years old on average, ranging from 23 to 62, with a median of 28). All participants have used the Google search engine, and a majority (63%) indicated that their estimated average search time was about 10 minutes. Overall, most participants were highly educated PhD students (14) with a high level of information search experience, which may not be representative of the broader population.

Participants were instructed to search for online information by assigned search tasks: two exploratory and up to twelve consecutive fact-finding tasks (Kotzyba et al., 2017, p. 90). The distinction between the fact-finding and exploratory tasks, which has been extensively studied in the context of information seeking (e.g., Schwerdt et al., 2021; Cole et al., 2013), was designed to examine the effect of task contexts on visual search behavior on the accuracy of decision outcome, i.e., whether participants can fulfill the requirements of information seeking. Search tasks are deemed simple when the answer can be obtained directly from the first search result or snippet on the SERP (search engine results page) using task-related queries. Conversely, a task is considered difficult if all snippets from the formulated queries lack information, requiring users to assess numerous documents or read a lengthy document to find the answer.

Each participant was allocated up to 40 minutes to perform both exploratory tasks (20 minutes per task) and, at most, 20 minutes to solve up to twelve factual tasks, using the Firefox web browser and the familiar Google search engine. A Latin-squared design was used to counter-balance the order effects of tasks. The users' self-perceived prior knowledge about the exploratory search task (how familiar were you with the topic? on a 5-point Likert scale) and fact-finding task (do you have any prior knowledge about the topic? yes or no), the answers, and the time spent within a search task were recorded. Overall, the participants had comparable knowledge about the assigned search tasks, with little information about the topics for both fact-finding and exploratory tasks.

The search interaction data was recorded and eye movements were captured by the Tobii X2-60 eye tracker, with a 60Hz data-sampling rate [1]. The software Tobii-Studio (version 3.4.2) was used for data processing and analysis. The gaze metrics included the broad categories of fixations, saccades, and pupil sizes, and they were summarized values within each search session. Absolute and relative saccadic angles were used to measure a user's change of visual search to another area. The absolute saccadic angles are measured regarding the horizontal axis, and the relative saccadic angles are measured regarding the last position of the previous fixation point (See Table 1 for definitions of gaze metrics.)

We focused on the selected gaze metrics as indicators of cognitive processes in decision-making because previous studies have demonstrated the relevance of these features toward visual exploration (making a saccade to a new object) or exploitation (revisiting an object that was previously fixated) in information seeking tasks (e.g., König et al., 2016; Spering, 2022; Wittek et al., 2016). The reanalysis has been concerned with fact-finding tasks only since it is a typical task-based reading activity. More detailed descriptions of the experimental design can be found in (Kotzyba et al., 2017).

## Data Analysis

We used a logarithmic cross-ratio analysis (Fleiss, Levin, & Paik, 2003) technique to determine if there is any significant relationship between gaze metrics and answer correctness. This technique was chosen because it is presumably resistant to sample selection bias. It has been used to analyze the relationships among individual differences, search behavior, and gaze behavior (Saracevic, Kantor, Chamis, & Trivison, 1988; Wittek et al., 2016).

Then we constructed mixed-effects models to determine the effects of search task difficulty, prior knowledge, and gaze metrics on search performance. Mixed effects distinguish between fixed effects that are due to experimental conditions and random effects that are due to individual differences in a sample. We chose the mixed-effects models because they are useful for the examination of the random effects of subjects and search tasks (Baayen, Davidson, & Bates, 2008). Despite the assigned search task difficulty by design, the search tasks represented in the study are still a sample of all the possible tasks.

We primarily used the lme4 package in R statistical computing software for model fitting (Bates, Mächler, Bolker, & Walker, 2015). We performed an automatic backward model selection of fixed and random parts of the model and the p-values were determined by Satterthwaite's degrees of freedom method, using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). The pseudo-R-squared values were determined by the procedure (Nakagawa & Schielzeth, 2013) and refinements of (Johnson, 2014), using the jtools package (Long, 2023). In addition to considering the fixed effects of task difficulty and user perception, the random effects of search task and user were considered in our full model construction and data fitting. Model assessments based on diagnostic checks for non-normality of residuals and outliers, distribution of random effects, and heteroscedasticity were conducted.

For example, concerning search task difficulty as fixed effects, the selected model was represented as

$$\mathrm{T}imeSpentSec \sim TaskDifficulty + (1 \mid Task) + (1 \mid User)$$

where random intercepts for task and user are specified with $(1 \mid Task)$ and $(1 \mid User)$ respectively.

# Results

The overall results suggest that search task difficulty dominates answer correctness, whereas prior knowledge does not have significant effects.

## Effect of search task difficulty on search performance

The descriptive analysis revealed that there was a statistically significant difference in the answer correctness across questions (one-way ANOVA, $F(11, 151) = 2.47$, $p = 0.007$, $p < .01$). However, there was no statistically significant difference in the answer correctness among the participants (one-way ANOVA, $F(18, 144) = 1.22$, $p = 0.25$, $p > .05$). Participants responded to 8.8 factual questions on average (min = 4, max = 12, SD = 2.6), and they exhibited a higher error rate when answering difficult questions (28.6% errors) compared to easy ones (10% errors).

Participants spent about one minute (57.54 secs) on average for each assigned search task of factual question-answering ($M = 57.54, SD = 36.39$) and 79.6% of all questions were answered correctly. Easy and hard search tasks were answered correctly at 88.7% and 70.7% respectively. Easy tasks of #5 and #6 were all answered correctly and the rest had correct answer rates above 80.0%. Hard tasks of #7, #10, and #11 were particularly challenging for participants, with an answer correctness rate below 65%, but tasks #8, #9, and #12 were answered correctly above 80.0% (See Kotzyba et al. (2017, p. 90) for task descriptions).

The overall results reveal that search task difficulty has a large effect on search performance, measured by time spent (See Figure 1). There was a very significant difference in the time spent in search task difficulty in easy versus hard tasks (one-way ANOVA, $F(1, 169) = 89.09$, $p < .001$). The mean difference in the time spent between easy and difficult tasks was 35.5 and 78.0 secs respectively. Our final model for the

[1] https://go.tobii.com/Tobii-Pro-X2-60-user-manual

Table 1: Summary of the relationship between gaze metrics and answer correctness (n gaze metrics = 162, n answer correctness = 162; statistical significance at 95%).

| | Definitions | $CutPoint(Mean)$ | *Odds Ratio* | *Log Odds* | *Stand. Error* | $t$-Stat *Value* |
|---|---|---|---|---|---|---|
| Gaze Metrics | | | | | | |
| FixNumb | Number of fixations | 338.99 | 0.46 | -0.77 | 0.39 | -1.96 |
| FixDurSum | Sum fixation duration | 71376.21 | 0.60 | -0.51 | 0.39 | -1.30 |
| FixDurMean | Mean of fixation durations | 213.12 | 1.64 | 0.50 | 0.39 | 1.28 |
| FixDurSD | SD of fixation durations | 166.56 | 1.99 | 0.69 | 0.40 | 1.70 |
| SacNumb | Number of saccades | 553.57 | 0.54 | -0.62 | 0.39 | -1.57 |
| SacDurSum | Sum of saccades durations | 32181.70 | 0.51 | -0.68 | 0.40 | -1.71 |
| SacDurMean | Mean of saccades durations | 56.29 | 0.57 | -0.56 | 0.40 | -1.40 |
| SacDurSD | SD of fixation durations | 48.06 | 0.64 | -0.45 | 0.40 | -1.12 |
| SaccadicAmplitudeMean | Mean distance between previous & current fixation location | 4.23 | 1.20 | 0.18 | 0.40 | 0.46 |
| SaccadicAmplitudeSD | SD of saccadic amplitude | 5.18 | 1.20 | 0.19 | 0.39 | 0.48 |
| AbsoluteSaccadicDirectionMean | Mean offset in degrees from previous to current fixation | 183.71 | 1.01 | 0.01 | 0.39 | 0.03 |
| AbsoluteSaccadicDirectionSD | SD of absolute saccadic directions | 118.87 | 0.38 | -0.97 | 0.40 | **-2.39** |
| RelativeSaccadicDirectionMean | Mean of relative saccadic directions | 181.17 | 1.37 | 0.31 | 0.39 | 0.80 |
| RelativeSaccadicDirectionSD | SD of relative saccadic directions | 106.26 | 0.31 | -1.18 | 0.42 | **-2.80** |
| PupilSizeMean | Mean of the sizes of pupil left and pupil right | 2.89 | 1.32 | 0.28 | 0.39 | 0.72 |
| PupilSizeSD | SD of pupil sizes | 0.15 | 0.77 | -0.26 | 0.39 | -0.65 |

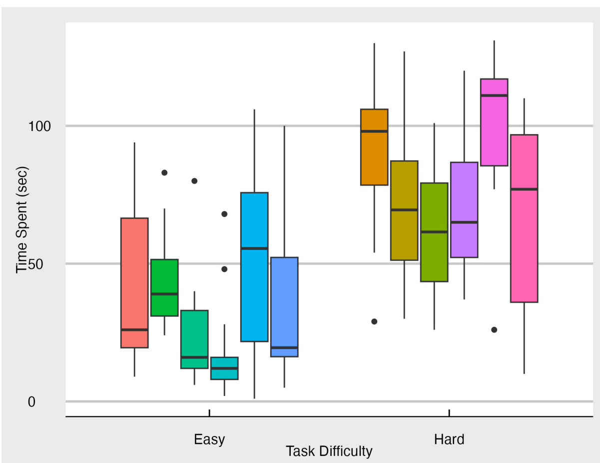


Figure 1: Boxplot of search task difficulty and time spent by search question.

Table 2: Effect of task difficulty on search performance by time spent and answer correctness.

| | **Time Spent** | **Answer Correctness** |
|---|---|---|
| TaskDifficulty | $42.34^{***}$ | $-0.18^{**}$ |
| | (7.78) | (0.08) |
| Constant | $37.05^{***}$ | $0.89^{***}$ |
| | (5.84) | (0.06) |
| N | 162 | 162 |
| Log Likelihood | $-765.68$ | $-81.58$ |
| AIC (Akaike Information Criterion) | 1541.40 | 173.15 |
| ICC (IntraClass Correlation) | 1556.80 | 188.59 |
| $R^2$ (fixed) | 0.34 | 0.05 |
| $R^2$ (total) | 0.49 | 0.13 |

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

effect of search task difficulty on time spent is a mixed-effects model that includes random intercepts for both the search task and the user. In contrast, our final model for answer correctness is a mixed-effects model that only includes search task difficulty as fixed effects.

So the results show that search task difficulty has a dominating effect on the correctness of answers, which confirms the validity of the instruments in the experiment (Table 2). Nonetheless, there are still individual differences in search performance by participants and specific tasks.

## Relationship between gaze metrics and answer correctness

Table 1 reveals that the gaze metrics of AbsoluteSaccadicDirectionSD (the SD of absolute saccadic direction) and RelativeSaccadicDirectionSD (the SD of relative saccadic direction) are correlated with answer correctness. Specifically, search sessions with a higher mean of AbsoluteSaccadicDirectionSD are more likely to have lower answer correctness by a factor of 0.38 (or 62%). Search sessions with a higher mean of RelativeSaccadicDirectionSD are more likely to have lower answer correctness by a factor of 0.31 (or 69%). Overall, the results reveal a significant relationship between the saccade trajectories and the answer correctness.

Other gaze metrics are not significantly correlated with the answer correctness, and the gaze metric of FixNumb (total number of fixations within a search session) is marginally significant. Given the role of prior knowledge in the control of saccade trajectories (Walker, McSorley, & Haggard, 2006), our further analysis has focused on the three metrics and their interactions with search task difficulty and the user's prior knowledge about the search tasks.

### Effect of gaze metrics and search task difficulty on answer correctness

As shown in Table 3, we specify mixed-effects models of gaze metrics (i.e., FixNumb, AbsoluteSaccadicDirectionSD, and RelativeSaccadicDirectionSD) and search task difficulty (as well as their interactional effects) as fixed effects and random intercepts for task and user. The results show that search task difficulty has a statistically significant effect on the answer correctness in where random intercepts for task and user are specified with $(1 \mid Task)$ and $(1 \mid User)$ respectively: Model 1: $FixNumb * TaskDifficulty + (1 \mid Task) + (1 \mid User)$ and Model 2: $AbsoluteSaccadicDirectionSD * TaskDifficulty + (1 \mid Task) + (1 \mid User)$. Since the AIC value of Model 2 is among the lowest of all the three models (AIC = 179.03), Model 2 is considered the best model. In Model 2, the interactional effect of AbsoluteSaccadicDirectionSD and TaskDifficulty is also statistically significant ($p < .05$). Since the fixed effect sizes of the Model 1 and Model 2 are 0.06 and 0.14 respectively, we can infer that the interactional effect of AbsoluteSaccadicDirectionSD and TaskDifficulty plays an important role in the explanatory power of Model 2. Overall, the results suggest the importance of the interactional effect of AbsoluteSaccadicDirectionSD and search task difficulty for predicting the answer correctness.

Table 3: Effect of gaze metrics and search task difficulty on answer correctness.

| | CorrectAnswer | | |
|---|---|---|---|
| | **Model 1** | **Model 2** | **Model 3** |
| FixNumb | −0.0005 | | |
| | (0.0003) | | |
| AbsoluteSaccadicDirectionSD | | −0.004 | |
| | | (0.01) | |
| RelativeSaccadicDirectionSD | | | −0.005 |
| | | | (0.004) |
| TaskDifficulty | −0.24** | 2.40** | 0.42 |
| | (0.11) | (1.10) | (0.66) |
| FixNumb:TaskDifficulty | 0.0004 | | |
| | (0.0003) | | |
| AbsoluteSaccadicDirectionSD:TaskDifficulty | | −0.02** | |
| | | (0.01) | |
| RelativeSaccadicDirectionSD:TaskDifficulty | | | −0.01 |
| | | | (0.01) |
| Constant | 0.97*** | 1.33 | 1.36*** |
| | (0.08) | (0.81) | (0.45) |
| N | 162 | 162 | 162 |
| Log Likelihood | −95.75 | −82.51 | −87.94 |
| AIC (Akaike Information Criterion) | 205.49 | 179.03 | 189.89 |
| ICC (IntraClass Correlation) | 0.08 | | 0.05 |
| $R^2$ (fixed) | 0.06 | 0.14 | 0.08 |
| $R^2$ (total) | 0.14 | 0.15 | 0.13 |

***p < .01; **p < .05; *p < .1

### Effect of gaze metrics and prior knowledge on answer correctness

Table 4 reveals that the user's prior knowledge about the search task has weak statistically significant effects on the answer correctness ($p < .01$). Based on the selected model 2 (AIC = 186.21), we can see that the user's prior knowledge about the search task and AbsoluteSaccadicDirectionSD and their interactional effects all have small effects on the answer's correctness and the fixed effect size is 0.10. Overall, the results suggest that prior knowledge does not have a statistically significant effect on the answer's correctness.

Table 4: Effect of gaze metrics and prior knowledge on answer correctness.

| | CorrectAnswer | | |
|---|---|---|---|
| | **Model 1** | **Model 2** | **Model 3** |
| FixNumb | 0.01* | | |
| | (0.004) | | |
| AbsoluteSaccadicDirectionSD | | −0.20* | |
| | | (0.11) | |
| RelativeSaccadicDirectionSD | | | −0.06* |
| | | | (0.03) |
| knownPrior [n] | 2.79** | −23.04* | −5.88 |
| | (1.40) | (13.59) | (3.70) |
| knownPrior [y] | 2.86* | −24.79* | −6.43 |
| | (1.46) | (14.72) | (4.33) |
| FixNumb:knownPrior [n] | −0.01* | | |
| | (0.004) | | |
| FixNumb:knownPrior [y] | −0.01 | | |
| | (0.004) | | |
| AbsoluteSaccadicDirectionSD:knownPrior [n] | | 0.19* | |
| | | (0.11) | |
| AbsoluteSaccadicDirectionSD:knownPrior [y] | | 0.20* | |
| | | (0.12) | |
| RelativeSaccadicDirectionSD:knownPrior [n] | | | 0.06* |
| | | | (0.03) |
| RelativeSaccadicDirectionSD:knownPrior [y] | | | 0.06 |
| | | | (0.04) |
| Constant | −1.94 | 25.75* | 7.44** |
| | (1.40) | (13.57) | (3.69) |
| N | 162 | 162 | 162 |
| Log Likelihood | −98.18 | −84.10 | −89.01 |
| AIC (Akaike Information Criterion) | 214.37 | 186.21 | 196.03 |
| ICC (IntraClass Correlation) | 0.08 | 0.03 | 0.08 |
| $R^2$ (fixed) | 0.04 | 0.10 | 0.05 |
| $R^2$ (total) | 0.12 | 0.12 | 0.13 |

***p < .01; **p < .05; *p < .1

### Effect of gaze metrics, prior knowledge and search task difficulty on answer correctness

Table 5 shows that search task difficulty and its interactional effects with AbsoluteSaccadicDirectionSD have statistically significant effects on the answer correctness ($p < .05$) in the selected model 2 (AIC = 192.63). The fixed effect size of the model is 0.16. Overall, the results suggest that search task difficulty has stronger effects on answer correctness than prior knowledge about search tasks. And the gaze metric of AbsoluteSaccadicDirectionSD and its interactional effect with search task difficulty can predict the answer correctness.

## Discussion

Search systems that can predict when a human may require assistance can increase the rate of successful interactions. We identified the impact of search task difficulty and the standard deviation of saccadic eye movement directions as predictors for search success in question-answering tasks.

Table 5: Effect of gaze metrics, prior knowledge, and task difficulty on answer correctness.

| | CorrectAnswer | | |
|---|---|---|---|
| | **Model 1** | **Model 2** | **Model 3** |
| FixNumb | 0.01* | | |
| | (0.004) | | |
| AbsoluteSaccadicDirectionSD | | −0.18* | |
| | | (0.11) | |
| RelativeSaccadicDirectionSD | | | −0.06* |
| | | | (0.03) |
| knownPrior [n] | 2.72* | −21.81 | −5.65 |
| | (1.39) | (13.32) | (3.70) |
| knownPrior [y] | 2.68* | −22.13 | −5.99 |
| | (1.45) | (14.47) | (4.37) |
| TaskDifficulty | −0.23** | 2.31** | 0.34 |
| | (0.11) | (1.11) | (0.67) |
| FixNumb:knownPrior [n] | −0.01* | | |
| | (0.004) | | |
| FixNumb:knownPrior [y] | −0.01 | | |
| | (0.004) | | |
| FixNumb:TaskDifficulty | 0.0004 | | |
| | (0.0003) | | |
| AbsoluteSaccadicDirectionSD:knownPrior [n] | | 0.18* | |
| | | (0.11) | |
| AbsoluteSaccadicDirectionSD:knownPrior [y] | | 0.18 | |
| | | (0.12) | |
| AbsoluteSaccadicDirectionSD:TaskDifficulty | | −0.02** | |
| | | (0.01) | |
| RelativeSaccadicDirectionSD:knownPrior [n] | | | 0.05 |
| | | | (0.03) |
| RelativeSaccadicDirectionSD:knownPrior [y] | | | 0.06 |
| | | | (0.04) |
| RelativeSaccadicDirectionSD:TaskDifficulty | | | −0.005 |
| | | | (0.01) |
| Constant | −1.75 | 23.12* | 7.01* |
| | (1.39) | (13.35) | (3.73) |
| N | 162 | 162 | 162 |
| Log Likelihood | −104.54 | −85.31 | −92.66 |
| AIC (Akaike Information Criterion) | 231.07 | 192.63 | 207.32 |
| ICC (IntraClass Correlation) | 0.07 | | 0.04 |
| $R^2$ (fixed) | 0.09 | 0.16 | 0.10 |
| $R^2$ (total) | 0.15 | 0.16 | 0.14 |

***p < .01; **p < .05; *p < .1

The actual cognitive processes underlying eye movements are complex if they go beyond a simple search task. As search tasks are performed, there can be partial reading, information processing, search for information, and even unconscious processes that can affect eye movements and mental workloads (J. Liu & Albright, 2018; Nocera, Camilli, & Terenzi, 2007). Further, the fixation pattern of human reading can be modeled by neural network-based attention with task-specific demands (Hahn & Keller, 2023). Goal-driven reading, like information-seeking reading (i.e., reading to answer questions), has been systematically different from ordinary reading: readers engage with the text strategically for optimizing cognitive resources, and eye movement patterns interact with task performance (Shubi & Berzak, 2023). Our work provides additional insights into the correlations between the features of eye movements and the search performance in answering fact-finding questions in information seeking.

From the perspective of adaptive search systems development, our findings can provide implications for computational research on the use of eye movement patterns and search interaction data to predict user characteristics (Cole et al., 2013; Conati et al., 2020; Toker, Conati, Steichen, & Carenini, 2013). In user interactions with visualization systems, research has also revealed the connection between the lower value of the standard deviation of absolute saccadic angles and search success (B. Fu & Steichen, 2019). However, many more features of eye movements have been identified as influential classification features, contributing to search success, including the standard deviation of relative saccadic angles and pupil dilation change. Despite the research findings regarding the sources of variability in saccadic eye movements in Neuroscience (van Beers, 2007) and cognitive modeling of gaze-based selection (Chen, Acharya, Oulasvirta, & Howes, 2021), future research needs to consider the relationships among the gaze features like the variability in saccadic eye movements, types of search tasks and specific features of user interfaces for modeling user visual and search behavior since they involve cognitive processes in decision making and task-based reading tasks. Deep learning methods can also be applied to characterize the reading patterns by considering the user's levels of expertise (e.g., Castner et al., 2022; Spiller et al., 2021).

The generalizability of the results can be enhanced by increasing the number of participants and search tasks. Since our analysis has focused on summarized user interaction data within a search session and measures of user-perceived search task difficulty were not collected, we were not able to investigate the changes in the user's state or level of knowledge when users were engaging with the search system. Future research also needs to consider how the adaptive system can be extended to deal with reasoning based on prior knowledge (Ragni & Johnson-Laird, 2020), complex search tasks, and natural language queries.

## Conclusion

In this study, we explored the representation of the user's knowledge state for the design of adaptive search systems by re-analyzing a user experiment dataset (n=19). Mixed-effects user models were constructed to specify which gaze metrics are reflective of user knowledge in task-based reading. Research findings suggest that there are statistically significant correlations between gaze metrics of the SD (standard deviation) of absolute/relative saccadic directions and search performance in fact-finding search tasks. The findings demonstrate the significance of the variability in saccadic eye movements in information-seeking, such as task-based reading. Importantly, search task difficulty and its interactional effect with the standard deviation of absolute saccadic directions have significantly contributed to the answer correctness. Nonetheless, the generalizability of the findings is limited by the small sample size. Implications for modeling user knowledge in information searching for the design of adaptive search user interfaces are discussed.

## Acknowledgments

Ying-Hsang Liu has been funded by the Visiting Scholar Program of Chemnitz University of Technology for established research partnerships and ERC project grant agreement ID: 818210.

## References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi: 10.1016/j.jml.2007.12.005

Bailly, G., Khamassi, M., & Girard, B. (2023). Computational model of the transition from novice to expert interaction techniques. *ACM Transactions on Computer-Human Interaction*, *30*(5), 1–33. doi: 10.1145/3505557

Barral, O., Lallé, S., & Conati, C. (2020). Understanding the effectiveness of adaptive guidance for narrative visualization. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '20)* (pp. 1–9). New York: ACM. doi: 10.1145/3377325.3377517

Barz, M., Bhatti, O. S., & Sonntag, D. (2022). Implicit estimation of paragraph relevance from eye movements. *Frontiers in Computer Science*, *3*, 1–13. doi: 10.3389/fcomp.2021.808507

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, *67*(1), 51. doi: 10.18637/jss.v067.i01http://arxiv.org/abs/1406.5823

Castner, N., Frankemolle, J., Keutel, C., Huettig, F., & Kasneci, E. (2022). LSTMs can distinguish dental expert saccade behavior with high "plaque-urracy". In *Proceedings of the Eye Tracking Research and Applications Symposium (ETRA '22)*. New York: ACM. doi: 10.1145/3517031.3529631

Chen, X., Acharya, A., Oulasvirta, A., & Howes, A. (2021). An adaptive model of gaze-based selection. *Proceedings of Conference on Human Factors in Computing Systems (CHI '21)*. doi: 10.1145/3411764.3445177

Cole, M. J., Gwizdka, J., Liu, C., Belkin, N. J., & Zhang, X. (2013). Inferring user knowledge level from eye movement patterns. *Information Processing \& Management*, *49*(5), 1075–1091. doi: 10.1016/j.ipm.2012.08.004

Conati, C., Lallé, S., Rahman, M. A., & Toker, D. (2020). Comparing and combining interaction data and eye-tracking data for the real-time prediction of user cognitive abilities in visualization tasks. *ACM Transactions on Interactive Intelligent Systems*, *10*(2), 1–41. doi: 10.1145/3301400

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Assessing significance in a fourfold table* (3rd ed.). Hoboken, NJ: John Wiley & Sons. doi: 10.1002/0471445428.ch3

Fu, B., & Steichen, B. (2019). Using behavior data to predict user success in ontology class mapping: An application of machine learning in interaction analysis. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)* (pp. 216–223). IEEE. doi: 10.1109/ICOSC.2019.8665670

Fu, B., & Steichen, B. (2022). Impending success or failure? An investigation of gaze-based user predictions during interaction with ontology visualizations. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. New York: ACM. doi: 10.1145/3531073.3531081

Fu, W. T. (2020). How cognitive computational models can improve information search. In W. T. Fu & H. van Oostendorp (Eds.), *Understanding and improving information search: A cognitive approach* (pp. 29–45). Cham: Springer International Publishing. doi: 10.1007/978-3-030-38825-6_3

Fu, W.-T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, *52*(3), 195–242.

González-Ibáñez, R., Esparza-Villamán, A., Vargas-Godoy, J. C., & Shah, C. (2019). A comparison of unimodal and multimodal models for implicit detection of relevance in interactive IR. *Journal of the Association for Information Science and Technology*, *70*(11), 1223–1235. doi: 10.1002/asi.24202

Gwizdka, J., & Dillon, A. (2020). Eye-tracking as a method for enhancing research on information search. In W. T. Fu & H. van Oostendorp (Eds.), *Understanding and improving information search: A cognitive approach* (pp. 161–181). Cham: Springer International Publishing. doi: 10.1007/978-3-030-38825-6_9

Hahn, M., & Keller, F. (2023). Modeling task effects in human reading with neural network-based attention. *Cognition*, *230*, 105289. doi: 10.1016/j.cognition.2022.105289

Hofmann, K., Mitra, B., Radlinski, F., & Shokouhi, M. (2014). An eye-tracking study of user interactions with Query Auto Completion. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM '14)* (pp. 549–558). New York: ACM. doi: 10.1145/2661829.2661922

Johnson, P. C. (2014). Extension of nakagawa & schielzeth's r2glmm to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944-946. doi: 10.1111/2041-210X.12225

Kim, J. (2005). Task difficulty in information searching behavior: Expected difficulty and experienced difficulty. In *Proceedings of the acm/ieee joint conference on digital libraries* (p. 383). Denver, CO: IEEE. doi: 10.1145/1065385.1065486

König, P., Wilming, N., Kietzmann, T. C., Ossandón, J. P., Onat, S., Ehinger, B. V., … Kaspar, K. (2016). Eye movements as a window to cognitive processes. *Journal of Eye Movement Research*, *9*(5). doi: 10.16910/jemr.9.5.3

Kotzyba, M., Gossen, T., Schwerdt, J., & Nürnberger, A. (2017). Exploration or fact-finding: Inferring user's search activity just in time. In *Proceedings of the Conference Human Information Interaction and Retrieval (CHIIR '17)* (pp. 87–96). New York: ACM. doi: 10.1145/3020165.3020180

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). {lmerTest} package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Liu, C., Liu, Y.-H., Gedeon, T., Zhao, Y., Wei, Y., & Yang, F. (2019). The effects of perceived chronic pressure and time constraint on information search behaviors and experience. *Information Processing & Management*, *56*(5), 1667–1679. doi: 10.1016/j.ipm.2019.04.004

Liu, J., & Albright, K. (2018). Exploring the roles of the unconscious in information search behaviors. *Journal of Librarianship and Information Science*, *50*(3), 332–342. doi: 10.1177/0961000618769984

Liu, J., Liu, C., Cole, M., Belkin, N. J., & Zhang, X. (2012). Exploring and predicting search task difficulty. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CKIM '12)* (pp. 1313–1322). New York: ACM. doi: 10.1145/2396761.2398434

Liu, Y.-H., Arnold, A., Dupont, G., Kobus, C., Lancelot, F., Granger, G., ... Matton, N. (2023). User evaluation of conversational agents for aerospace domain. *International Journal of Human-Computer Interaction*, 1–20. doi: 10.1080/10447318.2023.2239544

Liu, Y.-H., Thomas, P., Gedeon, T., & Rusnachenko, N. (2022). Search interfaces for biomedical searching: How do gaze, user perception, search behaviour and search performance relate? In *Proceedings of the Conference on Human Information Interaction and Retrieval (CHIIR '22)* (pp. 78–89). New York: ACM. doi: 10.1145/3498366.3505769

Long, J. A. (2023). jtools: Analysis and presentation of social scientific data [Computer software manual]. Retrieved from `https://cran.r-project.org/package=jtools` (R package version 2.2.2)

Malmaud, J., Levy, R., & Berzak, Y. (2020). Bridging information-seeking human gaze and machine reading comprehension. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL 2020)* (pp. 142–152). doi: 10.18653/v1/2020.conll-1.11

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142. doi: j.2041-210x.2012.00261.x

Nocera, F. D., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 271-285. doi: 10.1518/155534307X255627

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675. doi: 10.1037/0033-295x.106.4.643

Purcell, Z. A., Roberts, A. J., Handley, S. J., & Howarth, S. (2023). Eye movements, pupil dilation, and conflict detection in reasoning: Exploring the evidence for intuitive logic. *Cognitive Science*, *47*(6). doi: 10.1111/cogs.13293

Ragni, M., & Johnson-Laird, P. N. (2020). Reasoning about epistemic possibilities. *Acta Psychologica*, *208*, 103081. doi: 10.1016/j.actpsy.2020.103081

Raptis, G. E., Katsini, C., Avouris, N., Belk, M., Fidas, C., & Samaras, G. (2017). Using eye gaze data & visual activities to infer human cognitive styles: Method & feasibility studies. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17)* (pp. 164–173). doi: 10.1145/3079628.3079690

Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, *39*(3), 161–176. doi: 10.1002/(SICI)1097-4571(198805)39:3⟨161::AID-ASI2⟩3.0.CO;2-0

Schwerdt, J., Kotzyba, M., & Nurnberger, A. (2018). Inferring user's search activity using interaction logs and gaze data. In *Proceedings of the 2017 International Conference on Companion Technology* (pp. 1–6). doi: 10.1109/COMPANION.2017.8287075

Schwerdt, J., Kotzyba, M., & Nurnberger, A. (2021). Fact-finding or exploration: Characterizing reading strategies in user's search activities. *Proceedings of the IEEE International Conference on Human-Machine Systems, ICHMS 2021*, 1–6. doi: 10.1109/ICHMS53169.2021.9582460

Shubi, O., & Berzak, Y. (2023). Eye movements in information-seeking reading. *Proceedings of the Annual Meeting of the Cognitive Science*, *45*, Retrieved from https://escholarship.org/uc/item/6019k40d.

Spering, M. (2022). Eye movements as a window into decision-making. *Annual Review of Vision Science*, *8*(1), 427–448. doi: 10.1146/annurev-vision-100720-125029

Spiller, M., Liu, Y.-H., Hossain, M. Z., Gedeon, T., Geissler, J., & Nürnberger, A. (2021). Predicting visual search task success from eye gaze data as a basis for user-adaptive information visualization systems. *ACM Transactions on Interactive Intelligent Systems*, *11*(2), 1–25. doi: 10.1145/3446638

Steichen, B., Conati, C., & Carenini, G. (2014). Inferring visualization task properties, user performance, and user cognitive abilities from eye gaze data. *ACM Transactions on Interactive Intelligent Systems*, *4*(2), 1–29. doi: 10.1145/2633043

Thomas, P., Billerbeck, B., Craswell, N., & White, R. W. (2019). Investigating searchers' mental models to inform search explanations. *ACM Transactions on Information Systems (TOIS)*, *38*(1), 1–25.

Toker, D., Conati, C., Steichen, B., & Carenini, G. (2013). Individual user characteristics and information visualization: Connecting the dots through eye tracking. In *Proceedings of the Conference on Human Factors in*

*Computing Systems (CHI '13)* (pp. 295–304). doi: 10.1145/2470654.2470696

van Beers, R. J. (2007). The sources of variability in saccadic eye movements. *The Journal of Neuroscience*, *27*(33), 8757. doi: 10.1523/JNEUROSCI.2311-07.2007

Walker, R., McSorley, E., & Haggard, P. (2006). The control of saccade trajectories: Direction of curvature depends on prior knowledge of target location and saccade latency. *Perception & Psychophysics*, *68*(1), 129–138. doi: 10.3758/BF03193663

Wittek, P., Liu, Y. H., Darányi, S., Gedeon, T., & Lim, I. S. (2016). Risk and ambiguity in information seeking: Eye gaze patterns reveal contextual behavior in dealing with uncertainty. *Frontiers in Psychology*, *7*, 1790. doi: 10.3389/fpsyg.2016.01790