In a Hurry: How Time Constraints and the Presentation of Web Search Results Affect User Behaviour and Experience

Garrett Allen[1[0000-0003-4449-1510]], Mike Beijen[1[0000-0002-1773-3150]], David Maxwell[1[0000-0002-1701-5917]], and Ujwal Gadiraju[1[0000-0002-6189-6539]]

TU Delft, Delft, The Netherlands
g.m.allen@tudelft.nl, m.f.beijen@student.tudelft.nl, maxwelld90@acm.org, u.k.gadiraju@tudelft.nl

Abstract. Time constraints are commonplace in our daily lives. While literature in recent years from the *Information Retrieval (IR)* community has increased our understanding of the effects of time constraints on search, practical effects on search outcomes have rarely been evaluated. Little is known about how different search interfaces influence search outcomes and experiences in time-constrained search. This constitutes a knowledge gap that we aim to address in our work. Through a pre-registered 4×4 between-subjects crowdsourced user study, we investigate the influence of four different interfaces (*list view*, *grid-based view*, *absence of result snippets*, and *linear scanning pattern view*) on search outcomes and experiences under imposed time constraints (*no constraint* and constraints at *two*, *five*, and *eight* minutes). Results from our study indicate that user task performance is considerably affected by time constraints. In addition, as time constraints are tightened, a trade-off between querying rates and click depths arises. While no interaction effects between SERP interfaces and time constraints were ultimately found, findings from this study form an essential foundation for future work on how search result presentation may assist those searchers under strict time constraints.

Keywords: web search · time constraints · user interfaces · search behaviour analysis · task performance · user experience.

1 Introduction

Time constraints are commonplace in our lives, and can arise from various causes—such as a report deadline, or public transport disruption. Such constraints likely influence how we interact with the world around us, and can affect our judgement and decision-making abilities. Of course, the activity of searching for information is also subject to time constraints, which in turn can introduce pressure. Amongst other effects, experiencing these constraints and/or pressures has resulted in observations where individuals change their search strategies [23]. Differences in online search behaviour measures—such as query rates, dwell times, and the time spent examining documents—have also been observed [7, 8]. Such behavioural changes come with consequences. This could lead to a decrease in effectiveness, as demonstrated in a clinical decision-making study where the gained accuracy from using a web-based medical search system decreased from 32% to 6% as time pressure increased [38]. The literature on the effects of time constraints on web search task performance however provides limited insights. Therefore, directing research efforts into how a search engine may assist searchers under time pressure is justified. While search engines cannot mitigate the time pressure that (some) of their users may be experiencing, search engines *can* change the presentation of results on the *Search Engine Results Page (SERP)*—with designs to support time-constrained users. Various adaptations of SERPs (e.g., [18, 19, 28]) have been explored, yet the extent to which elements of the SERP may cater to individuals under time constraints in terms of task performance is a knowledge gap worthy of additional examination.

This work aims to investigate how different time constraints influence *task performance*, *user behaviour*, and *user experience*. We conducted a pre-registered, crowdsourced 4×4 between-subjects factorial design user study and examined the effects of four different SERP interfaces on user experience: ***(i)*** a standard *list-based view*; ***(ii)*** a *grid-based view*; ***(iii)*** a *list-based view without result snippets*; and ***(iv)*** a *linear scanning pattern view*. We explored whether an affinity for technology moderates the relationship between time constraints and task performance, over scenarios with no constraints present—and with time constraints of two, five, and eight minutes. We address the following research questions.

RQ1 *How do different time constraints influence task performance?*

RQ2 *How do different time constraints affect search behaviour?*

RQ3 *In what way are different UI designs susceptible to the effect of time constraints?*

RQ4 *What impact do different SERP interfaces have on user experience?*

RQ5 *To what extent does Affinity for Technology Interaction serve as a moderating variable for task performance?*

Our results show that stricter time constraints decrease task performance considerably—and as a result, some search behaviour measures are influenced. Exploratory results suggest that as time constraints tighten, a trade-off between query rate and click depth arises. While no significant interaction between SERP interfaces and time constraints emerged, this work lays an important foundation for future work on how search result presentation may assist time-constrained searchers. In line with open science principles, we publish all the materials, including questionnaires and data collected in this study on the *Open Science Framework*.[1]

[1] `https://osf.io/3wx42/`—last accessed 10th October, 2022.

2 Related Work

2.1 SERP Interfaces

Early work in the *noughties* by Dumais et al. [10] explored SERP layouts by grouping seven experimental interfaces into traditional list interfaces, such as those used by Google or Bing, and category interfaces[2]. The authors found that category-based interfaces were more effective than list interfaces in terms of search time. Additionally, inline summaries displayed underneath a search result's title, as are commonplace on contemporary SERPs, were found to be more useful than their hover summary counterparts. A hover summary was only provided when the user hovered over the related document's title. Through the addition of contextual information (including the *PageRank* score [32], overall and export popularity, and awards won), Schwarz and Morris [37] identified a significant improvement in credibility assessment of search results—increasing the accuracy to equal a user assessing the web page in its entirety.

Regarding user evaluation, Kammerer and Gerjets [19] investigated how the user interface may facilitate search result evaluations using a traditional list interface, and a tabular interface with results sorted into three categories: subjective, objective, and commercial. The findings showed that in the tabular interface, participants paid less attention to commercial results and more to objective results. While using tabular or "grid" interfaces [20], participants are motivated to select more trustworthy search results and to divide attention more equally between the results [18]. In a similar interface *without* categories, Joachims et al. [17] found that a greater number of results were inspected before the first result was clicked—thereby decreasing the role of the search result's ranking. This interface was provided as a means of mitigating the effects of position bias.

Investigating search result snippets, Clarke et al. [4] reported short or absent snippets have a negative effect on click-through rates. Maxwell et al. [28] found that longer snippets do not improve task performance, while Cutrell and Guan [9] found the opposite, adding information to search snippets improved task performance. In the same work, Cutrell and Guan [9] suggest designing SERP to focus the user's attention on the result metadata, i.e., the URL and title, to address the problem of long snippets being problematic for navigational search tasks. Therefore, in this work we explore whether using a SERP sans snippets within a time constrained environment and a SERP with snippets re-positioned to the right side of the other result data has an effect on user behaviour.

Various SERP have been designed with a specific goal in mind. Salmerón [36] for example compared a conventional, list-based design with a SERP consisting of a graphical overview presenting the relationship between web pages. Participants of this study were asked to examine the text of the result pages in both interfaces as if preparing for a test, after which they would be assessed on their comprehension. By signalling the relationships between web pages, an increased inter-text comprehension was provoked. Focusing on learning, Roy et al. [35]

[2] The category interface is akin to a search *directory*, much like how early commercial web search engines presented documents (like *Yahoo!*).

investigated the inclusion of active learning tools in the SERP interface, and found that note-taking increased the number of facts covered in post-task written essays by 34%—and highlighting resulted in 34% more subtopics covered. The SERP may also be used to reduce the *Search Engine Manipulation Effect (SEME)* [11] through low and high bias alerts [12]. In the same work, Epstein et al. [12] found that the SEME caused users to focus more on lower-ranked results. In contrast, Wu et al. [41] found adding an answer module to the SERP makes users focus more on the top results, but also improved user engagement and user satisfaction. User engagement also played an important role in the work by Foulds et al. [14], who concluded that users retrieved relevant documents faster—and with less frustration—when using a SERP without advertisements present. In response to users lacking familiarity with how search engines retrieve results and the findings of an investigation into user trust in search engines, Pan et al. [33] suggested the need for providing short explanations on SERP. Following this suggestion, Ramos and Eickhoff [34] experimented with search results explanations in terms of query term contribution bars per search result. Novin and Meyers [30] called for greater transparency in SERPs—establishing that search result explanations lead to increased transparency and search efficiency. There is a rich and varied body of work around SERP interfaces, yet little work addresses the *interplay between these interfaces and time pressured searchers.*

2.2 Time Pressure and Web Search

The effects of time on search is a well-researched topic. As previously mentioned, van der Vegt et al. [38] investigated the effects of time-constrained searching within the medical domain. They found that the accuracy of using an online medical search system decreased as time pressure increased. Crescenzi et al. [6] investigated the effects of time pressure through a crowdsourced user study, and found that participants who perceived to be under time pressure experienced lower search satisfaction and higher task difficulty. This work was later extended by Crescenzi et al. [8] with a comparable experimental setup, finding numerous significant effects of the time constraints on, amongst others, time pressure, task difficulty, and search performance satisfaction. In an earlier user study, Liu et al. [24] varied the presence of a time constraint. Participants subjected to no time constraints self-rated significantly higher pre-search confidence, better post-search performance, higher post-search familiarity with the topic, and greater knowledge acquisition. As for search strategy, Weenig and Maarleveld [39] found that participants adapt to time constraints via a more selective search strategy—and are not likely to accelerate their examination of items (i.e., spend less time on items). Liu and Wei [23] showed that when presented with time constraints, searchers move from a more *"economic"* search strategy to a more *"cautious"* approach. This was reflected by the fact that fewer results per query were viewed.

Delays in web search are also a topic of conversation. Crescenzi et al. [8] looked at the influence of system delays which caused participants to think the system was slower *only* when the delays were present in the second task performed. Of particular note, Arapakis et al. [1] found that the addition of query submission delays are noticed sooner by users of a fast search engine. The user's belief that the search engine would help them complete their search task decreased as this induced latency increased. In addition, Maxwell and Azzopardi [26] also experimented with delays to search. They found delays to both query submission and document download affected the behaviour of participants of their laboratory study regarding the time spent examining documents.

3 Study Design

In this study, we seek to understand the relationship between SERP interfaces and task performance, user behaviour, and user engagement while under time constraints. To this end, we conducted a 4×4 between-subject study, where the independent variables are the SERP interface and the time constraint. To identify proper time constraints, we turned to recent works involving search tasks with no time constraints to find insights on the search duration. We find that the average time spent on a search task ranges from a minimum of seven minutes to a maximum of ten minutes [16, 25, 42]. To ensure the constraints act as a source of time pressure, we use values of two, five, and eight minutes. We also include a no time constraint condition—acting as a baseline.

3.1 Search Task and SERP Interfaces

In a scenario-based search task, participants are instructed to imagine themselves as a journalist who has been assigned at the last minute to replace a colleague at an international forum discussion on a controversial topic, DNA cloning. Their job is to collect arguments connected to this topic in order to prepare for this role. Participants are allotted a training period with a practice topic and no time constraints to familiarize themselves with the SERP interface.

The participants perform their searches using a mock search engine, *BBT*. BBT displays the search frame on the left and the experiment frame on the right. The search results are presented according to which experimental condition the participant is randomly assigned. In the *list-view*, users are presented with the standard SERP interface that you see on mainstream search engines like Google or Bing. The *grid-view* presents the results in a 3×3 grid. In the *sa-view* participants are shown a SERP with no snippets included. The *ilsp-view* places the snippets to the right of the results in a traditional list view. No auto-completion is used during the search process to prevent inducing any biases. Screenshots of these interfaces can be seen on the companion page.[3]

For the duration of the task, participants can add additional arguments using an "Add argument" button below the text boxes in the experiment frame, or remove them by using the trash can icon at the top right of each argument text box. A "Review instructions" button above the text boxes allows participants to review task instructions at any time. Additionally, participants are provided

[3] `https://sites.google.com/view/icwe2023`

with a "Finalize" button to exit early if they feel they have collected enough information or if further search does not yield additional arguments.

A timer at the top of the experiment frame shows the remaining time if any. An alert in the form of a pop-up will be given when there is one minute remaining. It is made technically impossible to copy and paste arguments into the text boxes to encourage active involvement and prevent misbehaviour. The actual search is performed using the Bing web search API.[4] Participants are free to issue any queries and (re)visit any search result as they normally would. While making use of the search engine, search behaviour is recorded using LogUI [29]. To simulate a scenario as realistic as possible, there will be no 'review' part where participants can add, alter or remove arguments in the last-minute.

3.2 Metrics and Analysis

Once time is up, participants are taken automatically to the post-task questionnaire. The questionnaire is configured to assess participants for Affinity for Technology Interaction (ATI) [15], user engagement [31], and experienced time pressure. Both the pre-task and the post-task questionnaire contained an attention check. Due to the cognitive load of the task and to prevent distraction, no attention checks are shown during the search task itself.

Table 1: Scale description for the Interpretation of Data into Arguments (D-Intrp) and the Quality of Arguments (D-Qual) metrics.

Rating	Description
0	Facts contained within one argument with no association.
1	Association of two useful or detailed facts: $A \rightarrow B$.
2	Association of multiple useful or detailed facts: $A + B \rightarrow C$; $A \rightarrow B \rightarrow C$; $A \rightarrow B \therefore C$.

(a) D-Intrp metric.

Rating	Description
0	Facts within the argument are irrelevant to the subject; facts hold no useful information or advice.
1	Facts are generalized to the overall subject matter; facts hold little useful information or advice.
2	Facts fulfill the required information need and are useful.
3	A level of technical detail is given via at least one key term associated with the technology of the subject; statistics are given.

(b) D-Qual metric.

We utilized metrics for task performance, search behaviour, and user experience in our evaluation. With the aim of assessing the quality of arguments identified by the participants, we apply techniques introduced by Wilson and Wilson [40] based on Bloom's Taxonomy [3]. We measure the number of arguments submitted by each participant (*F-Argument*), a measure analogous to *F-Fact* in [40]. At the argument level, we measure how well the participants interpret data into arguments (*D-Intrp*) and the quality of the arguments (*D-Qual*).

[4] https://www.microsoft.com/en-us/bing/apis/bing-web-search-api

Details on the scales for each can be seen in Table 1a and Table 1b, respectively. D-Qual and D-Intrp are averaged across arguments into one final value per participant. At the participant level we measure the level of topic focus (*T-Depth*) by rating the coverage of subtopics on a scale of 0–3. The subtopics are: (*i*) benefits of cloning, (*ii*) safety considerations, (*iii*) ethical considerations, and (*iv*) drawbacks of cloning that are separate from safety or ethical considerations. When necessary, these metrics are adapted to fit the context of extracted arguments. Finally, we apply the ATI scale from [15] to understand to what extent ATI moderates the relationship between time constraints and task performance.

To measure search behaviour under time constraints, we capture result clicks in the form of the ranking position of the result, participant dwell time on the SERP, and perception of time pressure. We also capture whether or not participants opt to stop entering arguments early and determine the reliance on a single result for argument selection. User experience is measured using the User Experience Scale - Short Form by O'Brien et al. [31].

3.3 Participants

Participants for the user study were recruited using the online participant recruitment tool Prolific[5] and were rewarded at a rate of GBP 7.50/h for successfully completing the task, regardless of experimental condition. The required sample size of 431 participants was determined by conducting a power analysis for an ANCOVA using *G*Power* [13] with an effect size $f = 0.25$ (indicating a moderate effect), significance threshold $\alpha = 0.05/21 = 0.00238$ (due to the anticipated number of statistical tests), and a statistical power of $(1-\beta) = 0.8$. In total, we recruited 523 participants via Prolific, with 37 submissions being excluded for failed attention checks or low-effort responses.[6] The remaining 486 submissions were marked as valid.

Table 2: Analyses performed per research question.

Research Question	Independent Variables	Dependent Variables	Covariates
RQ1	Time contraint	Task performance	SERP interface, ATI, prior knowledge, topical interest, web search experience, perception of time pressure
RQ2	Time constraint	Search behaviour	SERP interface, prior knowledge, topical interest, web search experience, perception of time pressure
RQ3	Time constraint & SERP interface	Task performance & Search behaviour	Prior knowledge, topical interest, web search experience, perception of time pressure
RQ4	SERP interface	User experience	Time constraint, prior knowledge, topical interest, web search experience, perception of time pressure
RQ5	Time constraint	Task performance	SERP interface, ATI

[5] `https://www.prolific.co`—last accessed 10$^{\text{th}}$ October, 2022.

[6] Participants are considered *low-effort* if they did not complete the study or were inactive for 2+ minutes. We also exclude application and browser tab switches.

To determine answers for each research question, we analyze the main effects via an Analyses of Covariances (ANCOVAs). The covariates, independent, and dependent variables are described in Table 2. In the event of significant main effects, interaction effects are investigated in a post-hoc analysis using Tukey's honest significant difference test. We correct for Type-I error inflation with the Holm-Bonferroni correction (adjusted $\alpha = 0.05$). All statistical analysis was performed using SPSS 26.

4 Results

4.1 Descriptive Statistics

All participants recruited were native English speakers with a mean age of 25.65 years old (SD = 9.05). The youngest participant was 18 and the oldest 72 years old. The gender distribution is skewed toward males (male: 59.9%, female: 39.3%, other: 0.8%). The participants were balanced across the experimental conditions. The highest levels of education completed most prevalent in the participant sample are high school (37.7%), graduate degree (26.7%), and technical/community college (24.5%). Our participant pool consisted of practised searchers, with 92.2% indicating using a search engine to search the web at least once per day. Prior knowledge regarding the topic of DNA cloning varied but was at a moderately low level (mean = -0.55, SD = 1.20, scale: $[-2, 2]$). Prior interest in the topic was also moderate (mean = 0.42, SD = 1.15, scale: $[-2, 2]$). Participants possessed an adequate understanding of the search task, according to self-reported scores for the task definition (mean = 1.61, SD = 0.65, scale: $[-2, 2]$). Based on the reported levels of prior knowledge and prior interest, we argue that the search task adheres to the desired characteristics of exploratory search tasks and that participants understood what was asked of them [22].

A total of 21,763 arguments were submitted, with an average of 5.69 arguments per participant. Looking at topic focus (T-Depth), the best-covered subtopic was the benefits of cloning (45.9%) followed by ethical considerations (23.0%), safety considerations (11.3%), and drawbacks of cloning (8.1%). It is worth noting that a bias for arguments in favour of the topic is present among our participants. The D-Qual, D-Intrp, and T-Depth ratings were completed by the authors. Examples of how arguments were assessed can be found in Table 3. Participants rated the traditional `list-view` the highest in terms of user experience (mean = 3.61, SD = 0.51). Yet, only 27.6% of participants opted to look at a second SERP. Participants did not have a preference for the interface.

4.2 Statistical Analysis

To evaluate task performance, we conducted a set of one-way ANCOVA tests. Investigating the effect of the time constraints on the level of topic focus (T-Depth) revealed a significant effect ($F(3, 476) = 9.853$, $p < 0.001$, partial $\eta^2 = 0.058$). Post-hoc analysis revealed statistically significant differences at the $p <$

Table 3: Example assessments of D-Qual, D-Intrp, and subtopic.

Argument	**D-Qual**	**D-Intrp**	**Subtopic**
It can save animals from possibly extinction, or even species who were already extinct.	2	0	1
The negative of DNA cloning is that is can lead to in-breeding. This is because the same genotypes are reproducing.	3	1	4
One of the best advantages of DNA cloning is, it helps infertile couples to reproduce	1	0	1
Cons: DNA cloning present a lot of ethical and religious dilemmas	2	0	3
Reproductive cloning is controversial and may cause a lot of problems, since it creates two identical organism.	2	2	3

0.001 level between the 2 minutes time constraint (mean = 0.77) and 8 minutes time constraint (mean = 1.26). No significant effect of time constraint on quality of arguments (D-Qual) was found, ($F(3, 476) = 2.159$, $p = 0.092$, partial $\eta^2 = 0.013$). The effect of time constraints on interpretation of data into arguments (D-Intrp) was found to be statistically significant ($F(3, 476) = 10.46$, $p < 0.001$, partial $\eta^2 = 0.062$). A significant difference ($p < 0.001$) between the 2 minute time constraint (mean = 0.037) in relation to the 8 minute time constraint (mean = 0.301) and the no time constraint conditions (mean = 0.374) was also identified. The effect of time constraints on number of arguments (F-argument) was also found to be statistically significant ($F(3, 476) = 11.82$, $p < 0.001$, partial $\eta^2 = 0.069$) with the 2 minute time constraint (mean = 3.924) being significantly different to the 8 minute time constraint (mean = 6.974) and no time constraint conditions (mean = 6.337) with $p < 0.001$. This indicates that a general pattern where as time constraints tighten, task performance decreases.

A one-way ANCOVA revealed a statistically significant effect of time constraints on query rate ($F(3, 477) = 34.62$, $p < 0.001$, partial $\eta^2 = 0.179$). Significant differences at the $p < 0.001$ level existed between the 2 minute time constraint (mean = 0.829) in relation to the 5 minute time constraint (mean = 0.437), 8 minute time constraint (mean = 0.360), and no time constraint (mean = 0.267) condition. Thus, as time constraints tightened, the query rate increased. No statistically significant effects of time constraints on average length of queries was found ($F(3, 477) = 0.71$, $p = 0.545$, partial $\eta^2 = 0.004$).

Using two-way ANCOVA, no statistically significant interaction effect was found between time constraint and user interface with respect to topic focus (T-Depth; $F(9, 466) = 0.648$, $p = 0.756$, partial $\eta^2 = 0.012$), quality of arguments (D-Qual; $F(9, 466) = 1.608$, $p = 0.110$, partial $\eta^2 = 0.030$), interpretation of data into arguments (D-Intrp; $F(9, 466) = 1.653$, $p = 0.098$, partial $\eta^2 = 0.031$), number of arguments (F-Argument; $F(9, 466) = 0.627$, $p = 0.775$, partial $\eta^2 = 0.012$), query rate ($F(9, 466) = 1.268$, $p = 0.252$, partial $\eta^2 = 0.024$), and average length of queries ($F(9, 466) = 0.942$, $p = 0.488$, partial $\eta^2 = 0.018$).

A one-way ANCOVA revealed no significant effect of SERP interface on user experience ($F(3, 477) = 1.925$, $p = 0.125$, partial $\eta^2 = 0.012$).

The covariate ATI was also statistically insignificant with respect to topic focus (T-Depth; $F(1, 480) = 0.359$, $p = 0.549$, partial $\eta^2 = 0.001$), quality of arguments (D-Qual; $F(1, 480) = 0.682$, $p = 0.409$, partial $\eta^2 = 0.001$), interpretation of data into arguments (D-Intrp; $F(1, 480) = 2.185$, $p = 0.140$, partial $\eta^2 = 0.005$), and number of arguments (F-Argument; $F(3, 480) = 0.095$, $p = 0.758$, partial $\eta^2 = 0.000$).

5 Discussion

How do different time constraints influence task performance? With RQ1, we expected to find that stricter time constraints reduced task performance, and in fact found that 3 of the 4 task performance metrics (T-Depth, D-Intrp, F-Arguments) decreased significantly as time constraints tightened. These findings are in line with similar works reporting reduced task performance related to the presence of a time constraint [7]. Moreover, in line with [38], we find significant differences in task performance between different lengths of time constraints.

How do different time constraints affect search behaviour? For RQ2, we anticipated seeing various effects from time constraints on search behaviours. However, our results indicate only an increased query rate for participants in stricter time constraints. Our exploratory investigation suggests the increase in queries comes at the cost of click depth. This means that as time constraints tighten, participants rely more on fresh or reformulated queries rather than explore deeper in the results sets. Although the increased query rate corroborates prior research [7, 23, 5], other influences in behavioural metrics could not be established.

In what way are different UI designs susceptible to the effect of time constraints? Considering the susceptibility of SERP interfaces to the effects of time constraints (RQ3), the expected outcome was that of a noticeable interaction effect between the SERP interface and the task performance metrics. However, no such interaction effects were discerned in our study.

What impact do different SERP interfaces have on user experience? Past findings have demonstrated that various elements and their presentation on the SERP have been found to impact measures related to user experience, such as informativeness [27], satisfaction [2], and difficulty [21]. As such, we expected to find that the SERP interfaces affect user experience when time constraints are present (RQ4). Counter to this expectation, our analysis showed no significant impact of the interface on perceived user experience.

To what extent does Affinity for Technology Interaction serve as a moderating variable for task performance? Lastly, we questioned to what extent ATI moderates the relationship between time constraints and task performance (RQ5). ATI correlates with characteristics such as technology usage and learning success [15]. Motivated by such correlations, we sought to explore the relationship between ATI and task performance, as defined in Section 3.2. Our exploration uncovered no clear correlations.

5.1 Caveats and Limitations

This pre-registered study has some limitations despite the careful preparations and attention devoted to its design. The generalizability of this study is limited by the fact that only one topic was used in the search task. The decision to design the study in the presented form was based on the identified knowledge gaps in related literature, maximizing the potentially contribution. As a more technical limitation, using a proxy to serve search results such that all participants would see exactly the same page would be ideal. Presumably, the absence of a proxy mainly manifested itself in personalization differences in terms of ads and some web pages not being available in certain regions. Using a proxy, however, would increase response times, possibly affecting user experience. Also, no existing solutions were satisfactory in terms of cost or ease of use. Hence, the presence of a proxy did not outweigh its drawbacks.

6 Conclusions and Future Work

We presented a user study aiming to increase an understanding of the relationships between time constraints and SERP interfaces and task performance, user behaviour, and user experience. Participants were tasked with finding arguments in favor or against a controversial topic under a time constraint while using a mock search system. Task performance was evaluated using qualitative measures to examine whether participants understood the topic more deeply.

The results of the user study show that stricter time constraints reduce task performance. Additionally, tighter time constraints affect web search behaviour in terms of increased query rate. An exploratory finding suggests this comes at the cost of click depth, which increases as time constraints loosen. Using various SERP interfaces, the susceptibility to the effects of time constraints was investigated with no sensitivity found. Also, the various SERP interfaces used did not influence the user experience, implying that user experience was neither improved nor worsened. We did not find ATI to serve as a moderating variable in the relationship between time constraints and task performance. Exploratory findings have shown that participants who stopped the task early because they believed they had collected enough arguments had comparable task performance scores to those who used all the time available. Our results have strengthened the existing literature revolving around time constraints and SERP interfaces and made contributions to the knowledge gap on the interplay between time constraints and SERP interfaces.

Turning to future work, we propose that further research is needed to establish the generalizability of this work. The limiting factor of one topic used in this study leaves an open pathway for further studies, where more than one topic is considered. The experimental setup used in this work may serve as a foundation for such future studies. Additionally, the absence of effects on user experience due to SERP interfaces used in this work warrants further investigation. Further focused investigation is required to determine whether time constraints have a clear effect on users' perceived experience with a particular interface. Findings from such a study may inform ways to design SERP interfaces which can be used in support of searchers in other directions of the information retrieval field. Finally, this study was among one of the first to study various lengths of time constraints. Further research efforts devoted to closer examination of the sensitivity between these time constraints and task performance is required.

Acknowledgments. This work was partially supported by the SURF Cooperative, the Design@Scale Lab, and the TU Delft AI initiative.

References

1. Arapakis, I., Bai, X., Cambazoglu, B.B.: Impact of response latency on user behavior in web search. In: Geva, S., Trotman, A., Bruza, P., Clarke, C.L.A., Järvelin, K. (eds.) The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014. pp. 103–112. ACM, New York, NY, USA (2014)
2. Arapakis, I., Leiva, L.A., Cambazoglu, B.B.: Know your onions: Understanding the user experience with the knowledge module in web search. In: Bailey, J., Moffat, A., Aggarwal, C.C., de Rijke, M., Kumar, R., Murdock, V., Sellis, T.K., Yu, J.X. (eds.) Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015. pp. 1695–1698. ACM, New York, NY, USA (2015)
3. Bloom, B.S., Engelhart, M.B., Furst, E.J., Hill, W.H., Krathwohl, D.R.: Taxonomy of educational objectives. The classification of educational goals. Handbook 1: Cognitive domain. Longmans Green, New York (1956)
4. Clarke, C.L.A., Agichtein, E., Dumais, S., White, R.W.: The influence of caption features on clickthrough patterns in web search. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 135–142. SIGIR '07, Association for Computing Machinery, New York, NY, USA (2007)
5. Crescenzi, A., Capra, R., Choi, B., Li, Y.: Adaptation in information search and decision-making under time constraints. In: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval. p. 95–105. CHIIR '21, Association for Computing Machinery, New York, NY, USA (2021)
6. Crescenzi, A., Capra, R., Arguello, J.: Time pressure, user satisfaction and task difficulty. Proceedings of the American Society for Information Science and Technology **50**(1), 1–4 (2013)
7. Crescenzi, A., Kelly, D., Azzopardi, L.: Time pressure and system delays in information search. In: Baeza-Yates, R., Lalmas, M., Moffat, A., Ribeiro-Neto, B.A. (eds.) Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015. pp. 767–770. ACM, New York, NY, USA (2015)
8. Crescenzi, A., Kelly, D., Azzopardi, L.: Impacts of time constraints and system delays on user experience. In: Kelly, D., Capra, R., Belkin, N.J., Teevan, J., Vakkari, P. (eds.) Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval, CHIIR 2016, Carrboro, North Carolina, USA, March 13-17, 2016. pp. 141–150. ACM, New York, NY, USA (2016)
9. Cutrell, E., Guan, Z.: What are you looking for? an eye-tracking study of information usage in web search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 407–416. CHI '07, Association for Computing Machinery, New York, NY, USA (2007)
10. Dumais, S.T., Cutrell, E., Chen, H.: Optimizing search by showing results in context. In: Jacko, J.A., Sears, A. (eds.) Proceedings of the CHI 2001 Conference on

Human Factors in Computing Systems, Seattle, WA, USA, March 31 - April 5, 2001. pp. 277–284. ACM, New York, NY, USA (2001)
11. Epstein, R., Robertson, R.E.: The search engine manipulation effect (seme) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences **112**(33), E4512–E4521 (2015)
12. Epstein, R., Robertson, R.E., Lazer, D., Wilson, C.: Suppressing the search engine manipulation effect (SEME). Proc. ACM Hum. Comput. Interact. **1**(CSCW), 42:1–42:22 (2017)
13. Faul, F., Erdfelder, E., Lang, A.G., AG, B.: G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods Instruments & Computers **39**, 175–191 (05 2007)
14. Foulds, O., Azzopardi, L., Halvey, M.: Investigating the influence of ads on user search performance, behaviour, and experience during information seeking. In: Scholer, F., Thomas, P., Elsweiler, D., Joho, H., Kando, N., Smith, C. (eds.) CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021. pp. 107–117. ACM, New York, NY, USA (2021)
15. Franke, T., Attig, C., Wessel, D.: A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. Int. J. Hum. Comput. Interact. **35**(6), 456–467 (2019)
16. Gadiraju, U., Yu, R., Dietze, S., Holtz, P.: Analyzing knowledge gain of users in informational search sessions on the web. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. p. 2–11. CHIIR '18, Association for Computing Machinery, New York, NY, USA (2018)
17. Joachims, T., Granka, L.A., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting clickthrough data as implicit feedback. SIGIR Forum **51**(1), 4–11 (2017)
18. Kammerer, Y., Gerjets, P.: How the interface design influences users' spontaneous trustworthiness evaluations of web search results: Comparing a list and a grid interface. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. p. 299–306. ETRA '10, Association for Computing Machinery, New York, NY, USA (2010)
19. Kammerer, Y., Gerjets, P.: Effects of search interface and internet-specific epistemic beliefs on source evaluations during web search for medical information: an eye-tracking study. Behav. Inf. Technol. **31**(1), 83–97 (2012)
20. Kammerer, Y., Gerjets, P.: The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. Int. J. Hum. Comput. Interact. **30**(3), 177–191 (2014)
21. Kelly, D., Azzopardi, L.: How many results per page?: A study of SERP size, search behavior and user experience. In: Baeza-Yates, R., Lalmas, M., Moffat, A., Ribeiro-Neto, B.A. (eds.) Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015. pp. 183–192. ACM, New York, NY, USA (2015)
22. Kules, B., Capra, R.: Constructing exploratory tasks for a faceted search interface. In: The Workshop on Computer Interaction and Information Retrieval. pp. 18–21. HCIR '08, Microsoft Research, Redmond, WA, USA (2008)
23. Liu, C., Wei, Y.: The impacts of time constraint on users' search strategy during search process. In: Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology. ASIST '16, American Society for Information Science, USA (2016)

24. Liu, C., Yang, F., Zhao, Y., Jiang, Q., Zhang, L.: What does time constraint mean to information searchers? In: Elsweiler, D., Ludwig, B., Azzopardi, L., Wilson, M.L. (eds.) Fifth Information Interaction in Context Symposium, IIiX '14, Regensburg, Germany, August 26-29, 2014. pp. 227–230. ACM, New York, NY, USA (2014)
25. Liu, J., Mitsui, M., Belkin, N.J., Shah, C.: Task, information seeking intentions, and user behavior: Toward a multi-level understanding of web search. In: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval. p. 123–132. CHIIR '19, Association for Computing Machinery, New York, NY, USA (2019)
26. Maxwell, D., Azzopardi, L.: Stuck in traffic: how temporal delays affect search behaviour. In: Elsweiler, D., Ludwig, B., Azzopardi, L., Wilson, M.L. (eds.) Fifth Information Interaction in Context Symposium, IIiX '14, Regensburg, Germany, August 26-29, 2014. pp. 155–164. ACM, New York, NY, USA (2014)
27. Maxwell, D., Azzopardi, L., Järvelin, K., Keskustalo, H.: Searching and stopping: An analysis of stopping rules and strategies. In: Bailey, J., Moffat, A., Aggarwal, C.C., de Rijke, M., Kumar, R., Murdock, V., Sellis, T.K., Yu, J.X. (eds.) Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015. pp. 313–322. ACM, New York, NY, USA (2015)
28. Maxwell, D., Azzopardi, L., Moshfeghi, Y.: A study of snippet length and informativeness: Behaviour, performance and user experience. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 135–144. SIGIR '17, Association for Computing Machinery, New York, NY, USA (2017)
29. Maxwell, D., Hauff, C.: Logui: Contemporary logging infrastructure for web-based experiments. In: Hiemstra, D., Moens, M.F., Mothe, J., Perego, R., Potthast, M., Sebastiani, F. (eds.) Advances in Information Retrieval. pp. 525–530. Springer International Publishing, Cham (2021)
30. Novin, A., Meyers, E.M.: Making sense of conflicting science information: Exploring bias in the search engine result page. In: Nordlie, R., Pharo, N., Freund, L., Larsen, B., Russel, D. (eds.) Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR 2017, Oslo, Norway, March 7-11, 2017. pp. 175–184. ACM, New York, NY, USA (2017)
31. O'Brien, H.L., Cairns, P.A., Hall, M.: A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int. J. Hum. Comput. Stud. **112**, 28–39 (2018)
32. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1999)
33. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., Granka, L.A.: In google we trust: Users' decisions on rank, position, and relevance. J. Comput. Mediat. Commun. **12**(3), 801–823 (2007)
34. Ramos, J., Eickhoff, C.: Search result explanations improve efficiency and trust. In: Huang, J., Chang, Y., Cheng, X., Kamps, J., Murdock, V., Wen, J., Liu, Y. (eds.) Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020. pp. 1597–1600. ACM, New York, NY, USA (2020)
35. Roy, N., Torre, M.V., Gadiraju, U., Maxwell, D., Hauff, C.: Note the highlight: Incorporating active reading tools in a search as learning environment. In: Scholer, F., Thomas, P., Elsweiler, D., Joho, H., Kando, N., Smith, C. (eds.) CHIIR '21: ACM SIGIR Conference on Human Information Interaction and Retrieval, Canberra, ACT, Australia, March 14-19, 2021. pp. 229–238. ACM, New York, NY, USA (2021)

36. Salmerón, L., Gil, L., Bråten, I., Strømsø, H.I.: Comprehension effects of signalling relationships between documents in search engines. Comput. Hum. Behav. **26**(3), 419–426 (2010)
37. Schwarz, J., Morris, M.R.: Augmenting web pages and search results to support credibility assessment. In: Tan, D.S., Amershi, S., Begole, B., Kellogg, W.A., Tungare, M. (eds.) Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, BC, Canada, May 7-12, 2011. pp. 1245–1254. ACM, New York, NY, USA (2011)
38. van der Vegt, A., Zuccon, G., Koopman, B., Deacon, A.: How searching under time pressure impacts clinical decision making. Journal of the Medical Library Association : JMLA **108**(4), 564–573 (Oct 2020)
39. Weenig, M.W., Maarleveld, M.: The impact of time constraint on information search strategies in complex choice tasks. Journal of Economic Psychology **23**(6), 689–702 (2002)
40. Wilson, M.J., Wilson, M.L.: A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. J. Assoc. Inf. Sci. Technol. **64**(2), 291–306 (2013)
41. Wu, Z., Sanderson, M., Cambazoglu, B.B., Croft, W.B., Scholer, F.: Providing direct answers in search results: A study of user behavior. In: d'Aquin, M., Dietze, S., Hauff, C., Curry, E., Cudré-Mauroux, P. (eds.) CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020. pp. 1635–1644. ACM, New York, NY, USA (2020)
42. Xu, L., Zhou, X., Gadiraju, U.: Revealing the role of user moods in struggling search tasks. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 1249–1252. SIGIR'19, Association for Computing Machinery, New York, NY, USA (2019)