Incorporating Widget Positioning in Interaction Models of Search Behaviour

Nirmal Roy, Arthur Câmara, David Maxwell, Claudia Hauff
Delft University of Technology
Delft, The Netherlands
{n.roy,a.barbosacamara,d.m.maxwell,c.hauff}@tudelft.nl

ABSTRACT

Models developed to simulate user interactions with search interfaces typically do not consider the visual layout and presentation of a *Search Engine Results Page (SERP)*. In particular, the position and size of interface *widgets*—such as entity cards and query suggestions—are usually considered a negligible constant. In contrast, in this work, we investigate the impact of widget positioning on user behaviour. To this end, we focus on one specific widget: the *Query History Widget* ($\mathcal{QHW}$). It allows users to see (and thus reflect) on their recently issued queries. We build a novel simulation model based on *Search Economic Theory (SET)* that considers how users behave when faced with such a widget by incorporating its positioning on the SERP. We derive five hypotheses from our model and experimentally validate them based on user interaction data gathered for an ad-hoc search task, run across five different placements of the $\mathcal{QHW}$ on the SERP. We find partial support for three of the five hypotheses, and indeed observe that a widget's location has a significant impact on search behaviour.

1 INTRODUCTION

Economic theory, specifically *microeconomic theory*, assumes that an individual or *firm* will tend to maximise their profit—subject to budget or other constraints [48]. Microeconomic theory can also provide us with an intuitive means to model human-computer interactions [1]. Given a *demand* (that may arise from factors such as the nature of the context, the underlying task, or the system used), a *user* will exert *effort* to interact with the system by expending *internal resources* such as their working memory, attention, or energy. Users of a system will also incur a *cost* by expending *external resources* such as time, money, or physical effort (such as moving a mouse, or typing on a keyboard) [37]. In the context of *Information Retrieval (IR)*, interactions between the user and system may lead to *benefits* in terms of information obtained, or resolved information needs [5, 7]. Rational users looking to maximise profit from their interactions can do so by either maximising their benefit or by minimising their expended cost and effort—and thus subscribe to the *Principle of Least Effort* [55].

This research has been supported by *NWO* projects *SearchX* (639.022.722) and *Aspasia* (015.013.027).

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Assuming that searchers behave in a rational way (a reasonable assumption to make [3]), we can model their interactions with a search engine to obtain insights into the different decisions made during the interaction process. In turn, these insights can help us provide explanations as to *why* users behave in a certain way. Importantly, such a model allows us to generate *testable hypotheses* as to how user behaviour will likely change when interface designs are modified based on a cost/benefit analysis of interface elements. For example, a study by Azzopardi et al. [3] found partial support for the hypothesis that, *as the cost and/or effort of issuing a query increases, users of a search system will issue fewer queries and examine more documents per query.*

Traditionally, *Information Seeking and Retrieval (ISR)* models [8–10, 16, 22, 51] provide post-hoc explanations as to *what* happens during episodes of information seeking. While these models are undoubtedly useful, they have no predictive power: *we cannot employ them to learn what is likely to happen* in terms of user behaviour when changes are made to the retrieval system in question. This predictive power is necessary, for instance, in order to simulate the effects changes to the presentation of a *Search Engine Results Page (SERP)* have on user behaviour, without having to run many costly user studies. Ultimately, the goal here is to only run user studies on interface designs that have shown promise from prior simulations.

In contrast to aforementioned models of ISR, our work follows a recent line of research that focuses on building mathematical models based on *Search Economic Theory (SET)* [1, 2] which is inspired by microeconomic theory—or *Information Foraging Theory (IFT)* [42, 43]. These models allow us to relate changing costs (e.g., the cost of querying, or the cost of examining a search result snippet) to changing search behaviours. Prior works in this area have focused on how users interact with a ranked list [12, 38], their stopping behaviours [34, 52], the trade-off between querying and assessing [1–3], and browsing costs [6, 26]. In these aforementioned works, the SERP typically has a simple layout: the user can submit queries and assess documents. In addition, *interface components* (hereafter referred to as *widgets*) such as *Related Searches* are typically considered to be placed at a fixed position, and their specific position is not part of the formal model definition. However, contemporary SERPs are complex, and widgets can appear at various positions on the SERP as shown anecdotally in Table 1: there is no consensus on positioning or size of the *Related Searches* widget across web search engines. In addition, contemporary SERPs contain direct answers (leading to *good abandonment* [31, 52]), advertisements, and information cards—as well as result lists that integrate content from a number of different search verticals.

In our work, we focus on an aspect of individual *widgets* on a SERP that—as already mentioned—has so far been neglected in

Table 1: The placement of (as well as the number of) text columns, and the number of entries in the *Related Searches* widget across ten different web search engines. Results retrieved on May 2nd, 2021 for the query chess. *Placement* corresponds to the widget's position within the SERP.

Search Engine	Placement	#Columns	#Entries
bing.com	Bottom left	1	8
google.com	Bottom left	2	8
duckduckgo.com	Bottom left	1	8
yandex.com	–	–	0
ask.com	Upper right	1	12
yahoo.com	Bottom left	2	8
qwant.com	Upper right	1	8
baidu.com	Bottom left	3	9
ecosia.org	Bottom left	*No columns*	8
dogpile.com	Top left	1	8

mathematical representations of user interaction: the *positioning* of a given widget on a SERP. With this focus, we selected one specific SERP widget to provide an initial exploration of how to incorporate widget positioning into a SET-based model. Concretely, we focus on the *Query History Widget ($\mathcal{QHW}$)*, which is shown in Figure 3. It allows a user to view and thus reflect upon their recently issued queries during a search session. The widget is easy to understand for users, and involves only a small number of interactions—making it ideal as a first widget to employ for our exploration. Our main research question is therefore as follows.

RQ *How can we incorporate widget positioning information in a SET-based model?*

To answer this question, we first derive a SET-based model that considers a widget's positioning as an input variable. Based on our formal model, we derive five hypotheses as to the search behaviour users are likely to exhibit as the widget's positioning changes. Subsequently, in order to validate our model (and therefore also the inclusion of the positioning component in the model), we conduct a user study with $N = 120$ participants that each complete one ad-hoc retrieval task using a SERP with the $\mathcal{QHW}$—in one of five different positions[1]. We observe empirical evidence that provides partial support for three of our five hypotheses which shows that: *(i)* a widget's location influences search behaviour; and *(ii)* we are able to successfully create a formal interaction model, incorporating positioning, and mostly find evidence for our derived hypotheses.

2 MICROECONOMIC THEORY AND (INTERACTIVE) IR

Many models of ISR have been defined in the past [1, 8–10, 16, 18, 22, 43, 51]. They can generally be categorised into two groups: *descriptive* models [8–10, 16, 22, 51] and *formal* (mathematical) models [1, 18, 43]. The former provide us with intuitions and and a holistic view of a user's search behaviour (e.g., with the *Berrypicking model* [8], users *pick* through information patches—analogous to people collecting berries). While they provide us with explanations of why searchers behave in a particular manner, they do not allow us to *predict* how a user's search behaviour will change in response to changes to the SERP, the quality of the results, etc. For this step, formal models such as *Search Economic Theory (SET)* [1, 2], *Information Foraging Theory (IFT)* [42] or the *Interactive Probability Ranking Principle (iPRP)* [18] are required. With the increasing complexity of SERPs and the increasing amount of decisions users have to take during search episodes (and thus the ever growing number of experimental variants one would have to explore when exploring new interface variants), being able to rely on formal models to explore promising areas of the user interface design space is vital for cost-effective and efficient iterations of novel search interfaces.

[1]An overview of the different widget positions we employ for our experiments is shown in Figure 3.

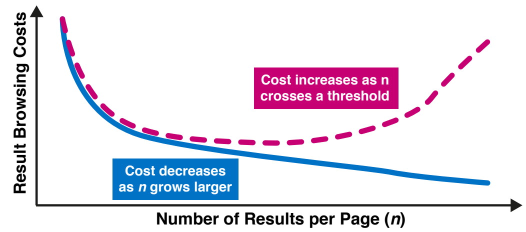

Figure 1: Example cost functions for two different SERP scenarios, adapted from Azzopardi and Zuccon [6]. Hypotheses can be derived from them, e.g. *the optimal number of results (n per page) to show to maximise a user's benefit in the* violet *scenario is at the cost function's global minimum.*

A key assumption of the listed formal models (which are related to each other as shown by Azzopardi and Zuccon [4]) is that users will modify their search behaviour to achieve the greatest possible net benefit from an interaction which is defined as the difference between the benefit of interaction and the cost of interaction. Thus, modelling the cost and benefit of interactions taking place on typical SERPs—and subsequently validating the designed models through user studies (or conversely finding that the proposed model is not sufficiently fine-grained enough to predict user behaviour well)—has been the focus of recent works in this area.

Specifically, Azzopardi and Zuccon [5] created user-oriented cost-benefit models to analyse a number of user decisions (including the length of the submitted query, the specificity of the query, the use of query suggestions vs. query reformulation, etc.) that are made during a search session—and at what point those decisions lead to maximum user benefit. The authors focus on model creation; the developed models are not empirically validated. In a similar vein, Azzopardi and Zuccon [6] developed a cost model to determine—for various *screen sizes*—the number of search result snippets that should be visible on the SERP, under the assumption that a user is looking for a document, and continues looking until that document is found. The developed formal model was initialised with hyperparameter values (such as the cost in seconds of typing out a character, or clicking a link) taken from the literature. Based on the developed cost functions (idealised examples of which are shown in Figure 1), several hypotheses were created—though their validation through a user study remained a point for future work. While this work already hinted at a distinction between desktop and mobile search (via the very different number of visible results in the viewport), Verma and Yilmaz [49] explicitly tackled this challenge and empirically determined (with 193 search sessions over $N = 25$ participants) to what extent existing user cost-benefit models are applicable (without change) to the mobile setting. The authors found that the parameters between desktop and mobile settings vary widely, and existing cost functions (with fixed hyperparameters, tuned to desktop search—and not adapted to the mobile setting) do not correlate very well with user satisfaction.

Using SET as their theoretical underpinning, Ong et al. [40] recently investigated the relationship between typing speed and search behaviour, both formally as well as empirically. While the authors did indeed observe a relationship between the two, they did find discrepancies between the observed user behaviours and those predicted by their model, conjecturing that their approximation of the model's query cost (by typing speed) does not capture all important aspects of the query cost component. A similar methodology was used by Maxwell and Azzopardi [33]. Here, the authors derived five different hypotheses about how temporal delays (both query response delays and document download delays) affect search behaviour. These hypotheses were derived from SET- and IFT-based models, respectively. Empirically (with $N = 48$ participants), three of the five hypotheses on user behaviours held.

Prior works *have* successfully employed formal models to derive testable hypotheses of search behaviours. To the best of our knowledge, none of the prior works have however considered the *position* of a user interface widget as important enough to include in the derived model. In our work, we focus on this very issue: *how does the position of a search interface widget impact the search behaviour our model predicts, and to what extent do those predictions hold when examining interaction data derived from a user study?*

3 CONSIDERING WIDGET POSITIONING

In this section, we first discuss—at a high level—how to incorporate the positioning of a widget within an interface in a SET-based model. We then introduce our implementation of the $\mathcal{QHW}$ in more detail, and present the cost functions for our specific widget use case. We conclude this section with a number of hypotheses we derive from our mathematical model regarding the influence of the $\mathcal{QHW}$ position on a user's search behaviour.

3.1 Positioning based on *Fitts' Law*

One way to consider the positioning of widgets within an interface—in a microeconomic cost model of interaction—is to *estimate the time it will take for a user to find the widget on an interface/SERP* from a given *starting position.* One way to approximate this is by using *Fitts' Law* [17]—an established, robust model of human psychomotor behaviour which has been has been frequently applied to computer and mobile interface design [25, 30, 46]. It states that the movement time for a user (moving their cursor on screen from a *source* to some *target*) is affected by the distance moved and the precision needed for such movement. The bigger and closer the target is, the easier it is to find and click. Shorter mouse movements are preferred, given that the object is large enough [46]. Therefore, given a search interface, the time taken to find a widget within a SERP is a function of its position and its size. In this work, Fitts' law is used as part of our SET-based user interaction model.

3.2 The Query History Widget (QHW)

Let us now turn our attention to $\mathcal{QHW}$, the interface widget that we developed the position-aware cost functions for. Shown in the callout in Figure 3, the $\mathcal{QHW}$ lists all previously issued queries

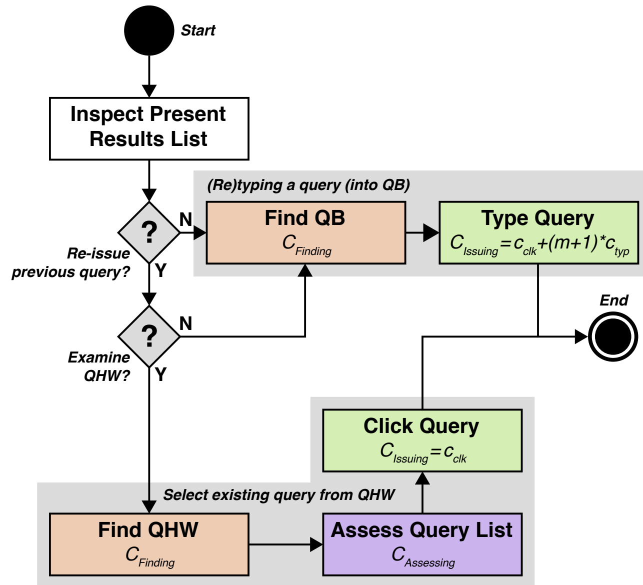

Figure 2: Flow chart of the modelled querying process. Before issuing each query, the user is presented with the choice of inspecting the $\mathcal{QHW}$ or typing in the query via the $\mathcal{QB}$. Associated costs are outlined in §3.2.1.

in a search session. Our model considers the following scenario. A user, after inspecting a retrieved list of documents presented on a SERP, decide to issue a different query. Do they: *(a)* reissue an old query (i.e., a query submitted earlier in the *same* search session, perhaps because they wish to find a document from earlier); or *(b)* issue a new query (i.e., a query not yet submitted in the *same* search session, potentially leading to a new set of documents)? If the user decides to issue a new query, they will head to the *query box* ($\mathcal{QB}$), and type the new query. If the user decides to reissue an old query, they must then decide whether to: *(a)* re-type the query in the $\mathcal{QB}$; or *(b)* scan the $\mathcal{QHW}$, find the old query, and click it. A flowchart of the process described is shown in Figure 2.

In this paper, we focus exclusively on the scenario where the user has *decided to reissue an old query.* We assume that the user knows they have issued this particular query in the past (a reasonable assumption, given that we only consider queries from a single search session), and expects to find it in the $\mathcal{QHW}$. We develop a formal model in order to predict and understand the scenario where they will choose to re-type this old query in the $\mathcal{QB}$, or when they will select it from the $\mathcal{QHW}$ instead—all *conditional on the position of the $\mathcal{QHW}$ on the SERP.* Note that this work does not focus on the reasoning behind re-issuing a query from earlier. We leave this for future work. Rather, we aim here to integrate positional information within a SET-based interaction model.

3.2.1 Specifying Costs. The total cost $C_{Reissuing}$ (in seconds) of reissuing an old query (that consists of m characters, and is listed at position q inside the $\mathcal{QHW}$) can be represented as three constituent components, as shown in the following equation.

$$C_{Reissuing} = C_{Finding} + C_{Assessing} + C_{Issuing} \tag{1}$$

$C_{Finding}$ is the cost of finding either the $\mathcal{QHW}$ or the $\mathcal{QB}$. We approximate the cost in terms of time taken (in seconds). According to Fitts' Law [17], the movement time of the mouse cursor from some starting position on a display to some target (in this case, either the $\mathcal{QHW}$ or the $\mathcal{QB}$) is equal to $a + b\log_2(\frac{D}{H} + 0.5)$ [11, 27], where D is the distance to the centre of the widget from the starting position of the cursor, H is the height of the widget (in *2D* interfaces, the smallest value from the target's height or width is considered [32]), and a, b are constants that are empirically determined. Intuitively, the further the widget is from the starting position, the more time it will take for the users to find the widget.

$C_{Assessing}$ is the cost of assessing a widget. For $\mathcal{QB}$, this cost is zero as users do not have to check a list of options. For $\mathcal{QHW}$, it involves two actions: *scrolling and checking.* For example, consider that a user wants to find the q^{th} query (our *target query*) in the $\mathcal{QHW}$. We associate a constant cost c_{scr} with scrolling over one query. Similarly, we associate a constant cost c_{chk} with checking whether a query is the *target query* or not. Given that the $\mathcal{QHW}$ displays t queries *above the fold* (e.g., in our experimental interface, as illustrated in Figure 3, we fixed $t = 4$), if $q \leq t$, then users do not incur any scrolling cost—and only the cost of checking to see if the query matches what they are seeking, or $q \times c_{chk}$. However, if $q > t$, users then have to scroll until the desired query is visible. This cost can be estimated by $(q - t) \times c_{scr} + q \times c_{chk}$, in line with [6].

$C_{Issuing}$ is the cost associated with entering the query. For $\mathcal{QB}$, it is the cost of typing the query of length m; this cost is $c_{clk} + (m + 1) \times c_{typ}$, where c_{clk} is the cost of clicking on the $\mathcal{QB}$, c_{typ} is the cost of typing one character and +1 is included to account for the pressing of ⏎. For $\mathcal{QHW}$, it is the cost c_{clk} of clicking on the desired query link.[2]

3.2.2 *When to use $\mathcal{QHW}$.*

Based on the previous section, we can now write the cost functions $C^{\mathcal{QB}}_{Reissuing}$ and $C^{\mathcal{QHW}}_{Reissuing}$: re-issuing an old query by typing into $\mathcal{QB}$, and by selecting a query from the $\mathcal{QHW}$, respectively. Based on our assumption of a rational user, we argue that a user will chose $\mathcal{QB}$ if the cost of using $\mathcal{QB}$ is less than the cost of using $\mathcal{QHW}$. For completeness, we present both cost functions in Equations 2 and 3 below, as well as a short definition of the corresponding symbols. For simplicity and neatness, we suppress the subscript from $C^{\mathcal{QB}}_{Reissuing}$ and $C^{\mathcal{QHW}}_{Reissuing}$ for now on, referring these costs simply as $C^{\mathcal{QB}}$ and $C^{\mathcal{QHW}}$, respectively. Rational users should choose $\mathcal{QB}$ over $\mathcal{QHW}$ if $C^{\mathcal{QB}} < C^{\mathcal{QHW}}$.

$$C^{\mathcal{QB}} = a + b\log_2(\frac{D_{QB}}{H_{QB}} + 0.5) + c_{clk} + (m+1) \cdot c_{typ} \qquad (2)$$

$$\begin{aligned} D_{QB} &= \text{Distance of } \mathcal{QB} \text{ from starting position} \\ H_{QB} &= \text{Height of query box (in pixels)} \\ m &= \text{Query length (in characters)} \\ c_{typ} &= \text{Cost of typing one character} \end{aligned}$$

$$C^{\mathcal{QHW}} = \begin{cases} a + b\log_2(\frac{D_{QHW}}{H_{QHW}} + 0.5) + q \cdot c_{chk} + c_{clk}, & \text{if } q \leq t \\ a + b\log_2(\frac{D_{QHW}}{H_{QHW}} + 0.5) + (q - t)c_{scr} + q \cdot c_{chk} + c_{clk}, & \text{if } q > t \end{cases} \qquad (3)$$

$$\begin{aligned} D_{QHW} &= \text{Distance of } \mathcal{QHW} \text{ from starting position} \\ H_{QHW} &= \text{Height of } \mathcal{QHW} \text{ (in pixels)} \\ q &= \text{Position of the target query in } \mathcal{QHW} \\ c_{chk} &= \text{Cost of checking a query in } \mathcal{QHW} \\ c_{scr} &= \text{Cost of scrolling over a query in } \mathcal{QHW} \\ c_{clk} &= \text{Cost of clicking a hyperlink in } \mathcal{QHW} \end{aligned}$$

3.2.3 *Constants.*

In our model, the above inequality depends not only on the positioning of $\mathcal{QHW}$, but also on the value of a few constants. These are: the cost of clicking(c_{clk}); scrolling(c_{src}); typing (c_{typ}); checking(c_{chk}) queries; the sizes of both $\mathcal{QB}$(H_{QB}) and $\mathcal{QHW}$ (H_QHW); the number of queries above the fold (t); the distance from the bottom of the screen to $\mathcal{QB}$ (D_{QB}); and the considered starting point of the cursor. In order to derive meaningful hypotheses from our inequality and use the model to predict actual user behaviour, we need to provide meaningful estimates of these constants. We can either estimate them directly from the interaction logs we collect in our user study, or fix their values based on studies reported in the literature in line with [4, 6]. For example, the typical values of pertaining are shown in Table 2 where we take c_{typ}, c_{scr}, c_{clk} and the hyperparameter b from the literature. We note that a—defined in Equations 2 and 3—is cancelled in our comparison, and thus is ignored. In order to make use of the model, we also need to define certain other constants like distance of $\mathcal{QB}$, or the height of $\mathcal{QHW}$, etc.—which we also report in Table 2. We need these precise values to predict real world behavior by calculating the exact cost of each decision. We leave this as future work. In this paper, we focus on using the general intuition behind the model equations to derive hypotheses of user interaction.

4 HYPOTHESES

Having defined our model in Equations 2 and 3—along with all associated constants, we now derive five hypotheses pertaining to the query issuing behaviour that the model describes, and how position can influence search behaviours.

H1 *As the length of query q to be reissued increases, a user will be more likely to reissue the query via $\mathcal{QHW}$.*

This first hypothesis follows from Equation 2. As m (the length of query q in characters) increases, $C^{\mathcal{QB}}$ increases. At the same time, m does not influence $C^{\mathcal{QHW}}$.

H2 *If the number of queries to check in $\mathcal{QHW}$ increases, a user's likelihood of using $\mathcal{QHW}$ increases as its distance to the starting point decreases.*

In Equation 3 we see that, if q increases, D_{QHW} has to decrease to keep the overall cost of using $\mathcal{QHW}$ lower than that of $\mathcal{QB}$.

H3 *The lower the distance of the $\mathcal{QHW}$ to the starting point, the more likely users are to use it.*

[2] While we have not yet described our implemented $\mathcal{QHW}$ widget in detail, we note that each old query is represented as a hyperlink; clicking a hyperlink reissues the query and displays the results for it on the SERP.

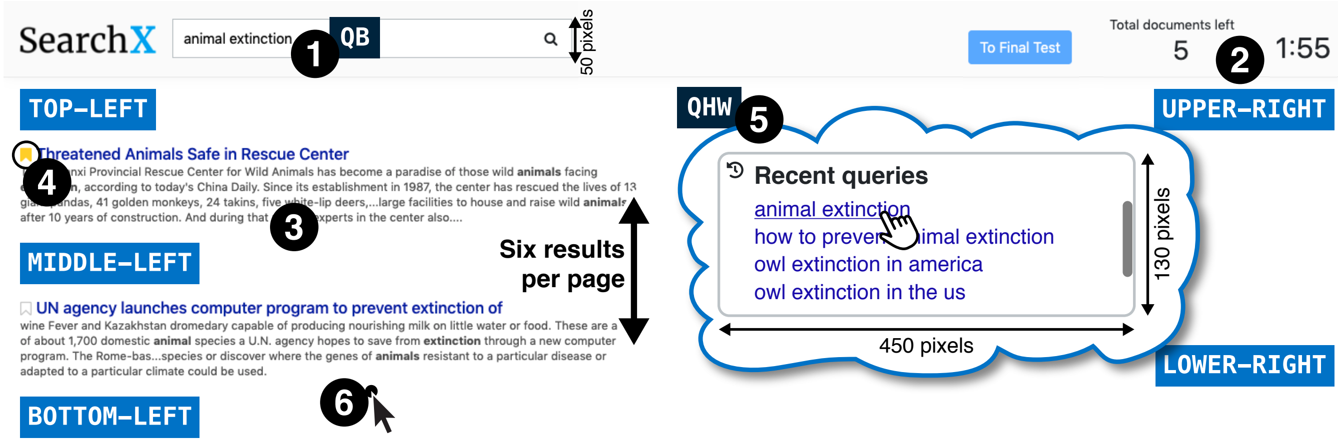

Figure 3: The SearchX search interface used for this study. Note the inclusion of the $\mathcal{QHW}$ in the callout—this was positioned in one of the areas as shown with blue boxes. Refer to §5.2 for information on the circled interface components.

Table 2: Overview of the model's constants and values used.

Constant	Value
Taken from the literature	
c_{typ}	0.28 [3, 11]
c_{scr}	0.1 [7]
c_{clk}	0.2 [3, 11]
b	0.1 [11, 27]
Defined for our experiments	
D_{QB}	1000px
H_{QB}	50px
H_{QHW}	130px
t	4
c_{chk}	0.25
Cursor starting position	End of search result list at bottom of screen, 6 from Figure 3

This follows from Equation 3 where, everything else being constant, the cost of reissuing a query is lowest when $D_{QHW} = 0$.

H4 *Users who type more slowly are more likely to use the $\mathcal{QHW}$ irrespective of where it is located.*

In §3.2.3, we provided fixed estimates for various constants in our model. One of those estimates is the cost of typing a character. Since the typing speed of users might vary to a considerable degree [40], the typing cost should be subject to further scrutiny. A user with slower typing will have a higher cost of typing queries, which will likely affect what widget they will use to reissue a query. For slow typing, c_{typ} is high, and $C^{\mathcal{QB}}$ becomes higher than $C^{\mathcal{QHW}}$ for all reasonable values of D_{QHW} and q. Hence, with slow typing, the positioning of $\mathcal{QHW}$ is less crucial—or how many queries are present in it, as a user is more likely to use $\mathcal{QHW}$ anyway.

H5 *A user's attention follows a F-shaped gaze pattern.*

This pattern has been observed on heterogeneous SERPs in the past [15] and should be reflected in the amount of attention users pay to the $\mathcal{QHW}$ in different positions. Specifically, the interface with $\mathcal{QHW}$ in the top right corner of the screen is likely to receive more attention than $\mathcal{QHW}$ positioned at the bottom right corner. Similarly, $\mathcal{QHW}$ in the bottom left corner is likely to receive less attention than $\mathcal{QHW}$ in the top left part of the screen.

5 USER STUDY DESIGN

In order to examine whether there is support for our hypotheses, we conducted a between-subjects user study. Participants were presented with a SERP that was complemented with the $\mathcal{QHW}$ in different positions, depending on the condition they were assigned.

5.1 System, Corpus, Topic and Task

For our user study we employed SearchX [44], a modular, open-source search framework which provides quality control features for crowdsourcing experiments. We integrated the LogUI framework [36] into SearchX to allow us to accurately capture all keyboard events and mouse events (including hovers and clicks) over $\mathcal{QB}$, $\mathcal{QHW}$, and results.

SearchX was configured to use the *TREC AQUAINT* corpus. The corpus consists of over one million newspaper articles from the period 1996 – 2000. Articles were gathered from three newswires: the *Associated Press (AP)*, the *New York Times (NYT)*, and *Xinhua*. Using a traditional test collection provided us with the ability to easily evaluate the performance of participants where required. We index the collection using *Indri*, and use its own snippet generator for the summaries presented to participants. We employed Indri's Dirichlet prior smoothing model (with $\mu = 2500$).

We used the `wildlife extinction` topic (topic number 347) from the *TREC 2005 Robust Track* [50]. A total of 165 relevant documents were identified by TREC assessors for this topic within AQUAINT. This topic was selected it as it has been successfully employed in prior user studies [3, 33]; the topic remains relevant to this day, and is likely to be of some interest to our participants.

We instructed our participants to identify documents that they perceived to be relevant to the TREC topic description that we provided to them. We primed our participants by asking them to imagine that they were to write an essay on the topic, and would use the identified documents as potential references at a later time.

5.2 Interface and Incentives

Our search interface is presented in Figure 3. It contains: the standard query box $\mathcal{QB}$ (without autocompletion features) ❶; a task timer and a bookmarked-documents counter ❷; six search *results per page (RPP)* ❸; functionality to mark documents in the form of a toggle icon ❹; and $\mathcal{QHW}$ ❺.

As we were looking to incentivise participants to reissue existing queries, special considerations needed to be made to this effect—along with considering that the search interface used should be kept simple to avoid any undue attention given to components that were not considered by our model defined in §3.2. We evaluated our incentives in a small pilot study before deploying them to our study participants. Results from the pilot study are not included in our final analysis.

Participants were instructed that they could mark no more than *six* documents at a time. The marked documents counter helped participants to keep track of their number of marked documents. The idea behind this was that a strict limit on how many documents could be marked would incentivise participants when issuing queries later on in their search session (either via $\mathcal{QB}$ or $\mathcal{QHW}$) to unmark previously marked documents (by toggling the icon)—and mark new ones that they perceived to be more promising. Participants were incentivised further by the potential for a bonus being awarded to the top six participants who achieved a high accuracy.

Before the study commenced, the participant's screen resolution was checked—a resolution check ensured that the resolution of the browser was 1920×1080 or greater. This resolution was found to show (with a high degree of certainty) that the entire search interface could fit on the participant's screen without the need for scrolling, meaning all six RPP were displayed, with none hidden *below the fold*. It also helped us to estimate the value of D_{QB} as presented in Table 2. There is also no pagination enabled on the SERP. These are due to the fact that our model does not include *page* scrolling[3] or pagination, factors that could alter user behaviour. To this end, we also removed any hyperlinks to documents. To compensate, we increased the number of lines for each summary snippet from the established two to four. While longer snippets have been shown to increase confidence in decisions of relevance at the expense of accuracy [35], it was decided that additional surrogate text in this instance would help participants in judging documents without access to the full text.

5.3 Operationalising the QHW

We operationalised the $\mathcal{QHW}$ as shown in the callout in Figure 3. The widget measures 450×130 pixels. At the top of the widget is the **`Recent queries`** title. Each query issued by the participant during the study is then prepended to the list shown in the lower portion of the widget. Queries are listed in reverse chronological order, with the most recently issued query appearing at the top.

As the $\mathcal{QHW}$ has a fixed width and height on the SERP, the widget could display at most four queries at a time, matching $t = 4$ as outlined in §3.2.1. Participants who wished to see more queries could scroll using their trackpad or mouse wheel to reveal older queries. All queries listed in the $\mathcal{QHW}$ were displayed as the standard blue hyperlink text—which underlines when hovered over—to provide a *proximal cue* [14] that they were hyperlinks that could be clicked. A click on the listed query then submits the listed query to the search engine, and displays the top six ranked documents.

In terms of positioning within the SERP, we trialled five different positions which are demonstrated in Figure 3 with blue boxes. Anecdotal evidence as presented in Table 1 suggests that there is no clearly defined position for widgets on a SERP (beyond the search results and entity cards), and thus we evaluated the major positions. Each of our five $\mathcal{QHW}$ positions (three on the left rail and two on the right) are represented in our user study as a unique condition.

`TOP-LEFT` Positioned at the top left, before the first result.

`MIDDLE-LEFT` Positioned on the left rail, below the third result.

`BOTTOM-LEFT` Positioned at the bottom left, immediately after the sixth and final result.

`UPPER-RIGHT` On the right rail, this condition positioned $\mathcal{QHW}$ underneath the clock; it is top-aligned with the first result. This position would be analogous to where an entity card sits on a contemporary web search engine's SERP.

`LOWER-RIGHT` On the right rail, this condition positioned $\mathcal{QHW}$ under the clock; it is aligned at the bottom with the last result.

5.4 Post-Task Survey

Inspired by the *User Experience Questionnaire (UEQ-S)* [20, 29], we asked participants five questions after the completion of the search task which. Questions explored the usage experience of the $\mathcal{QHW}$. All questions were answered using a 7-point Likert scale, considering negative to positive responses. For example, to understand to what extent a widget positioning was unexpected for the participant, we ask *"What did you think about the position of the query history widget?"*, with the scale ranging from *unexpected (1)* to *expected (7)*. Additionally, we ask about the support, ease of use, efficiency & clarity of the widget. Participants also received an open question for general comments and feedback about the interface.

5.5 Crowdsourced Participants

Participants for our study were recruited from *Prolific*, a crowdsourcing platform which has been shown to be an effective choice for complex and time-consuming *Interactive Information Retrieval (IIR)* experiments [53]. In order to obtain high-quality and reliable data, we imposed the following constraints: *(i)* participants needed to have at least 100 prior Prolific submissions; *(ii)* have an approval rate of 95% or higher; and *(iii)* have native proficiency in English. The complete study took approximately fifteen minutes, which included the minimum search time of 10 minutes. For their time, participants were compensated at the rate of GBP£8.00 per hour.

Overall, a total of 125 participants took part in our study. From this total, we had to reject five as they did not comply with our quality checks[4]. Our final cohort of 120 participants included 40 female and 80 males ones, with a reported average age of 35 years (youngest 18; oldest 77).

[3] It does however consider scrolling costs *within* $\mathcal{QHW}$.

[4] Our quality checks required that participants did not change the browser tab more than three times during the study, issued at least two queries, and marked at least two documents during their search session.

6 RESULTS

We now discuss the empirical validation of each of our five hypotheses which were introduced in §4. Recall that our research question asks whether widget positioning information can be meaningfully incorporated in a SET-based model.

A comparison of the main search behaviour indicators across conditions is shown in Table 3. On average, participants issued 12 queries (28 characters long)—and marked six documents, hitting the imposed limit). 114 participants reissued 5 queries on average, while six did not reissue any queries (either via $\mathcal{QB}$ or $\mathcal{QHW}$). On average our participants spent 12 minutes on the search task. We collected, on average, 2148 log events per participant.

Additionally, we also measured how the participants behaved regarding marking documents. On average, participants marked 2.60 relevant documents during their session, and 5.10 non-relevant documents. As expected, participants also unmarked documents over their session, indicating that they were actually reflecting on what they had marked. On average, participants unmarked 1.70 documents, where 1.15 of these were non-relevant.

The results of our post-task survey indicate that our interface was easy to use (Table 3, row **XV**: on average a score above 5 on a 7-point Likert scale), and the purpose of the $\mathcal{QHW}$ was clear (row **XVII**: on average a score above 5). Apart from `MIDDLE-LEFT`, which received a comparably low and significantly worse *expected position* score than almost all other variants (the only exception being `LOWER-RIGHT`), the $\mathcal{QHW}$ variants were positioned at somewhat expected locations (Table 3, row **XIII**: on average a score above 4 on the 7-point Likert scale).

These numbers indicate that our task design (which encouraged the reissuing of queries) was successful. Finally, we point the reader to Figure 3 for examples of actual queries our participants submitted (as visible in the $\mathcal{QHW}$ callout). We also considered ❻ from Figure 3 as the expected starting point where the cursor is positioned after they have scanned the search results. From this location, they move the cursor to $\mathcal{QHW}$ or $\mathcal{QB}$ to (re)issue queries. We argue that it is reasonable to expect individuals to examine all six results on the SERP before moving on. Coupled with the known correlation of eye gaze and cursor positioning on the screen [13], this assumption allows us to make estimations of D_{QB} and D_{QHW} in Equations 2 and 3.

6.1 H1: Query Length

Hypothesis **H1** states that as the length of some query to reissue increases, the likelihood of reissuing the query via $\mathcal{QHW}$—independent of the widget's position—increases. To investigate **H1**, we consider all 590 reissued queries across all participants and conditions. Queries were partitioned into four groups (with boundaries at the 25$^{\text{th}}$/50$^{\text{th}}$/75$^{\text{th}}$ percentiles), according to their length m in characters—$[1, 18]$, $[19, 25]$, $[26, 33]$ and $[34, \infty]$—and determined the fraction of queries reissued via $\mathcal{QB}$ and $\mathcal{QHW}$. Results are shown in Figure 4. We find that for the shortest reissued queries (with $m < 19$), 74% of queries are reissued via $\mathcal{QHW}$, while this percentage rises to 94% for the longest ($m > 33$) queries. This trend provides support for **H1**: participants prefer to use $\mathcal{QHW}$ for reissuing queries, and do so with a greater likelihood as query length increases. In order to observe if the trend is significant, we sampled one random reissued query from each participant to make the observations in the four query groups independent. A Chi-square test revealed that there is significant difference across the four query groups ($\chi^2(3, N = 114) = 10.58, p = 0.01$). Post-hoc tests showed that there were significant differences in the number of queries issued via $\mathcal{QHW}$ between queries belonging to the 3 larger size groups when compared to the group representing smaller queries. However, there was no significant difference among the three groups representing larger queries. We therefore find partial support to our hypothesis that people are more likely to use the $\mathcal{QHW}$ to reissue queries as query length increases.

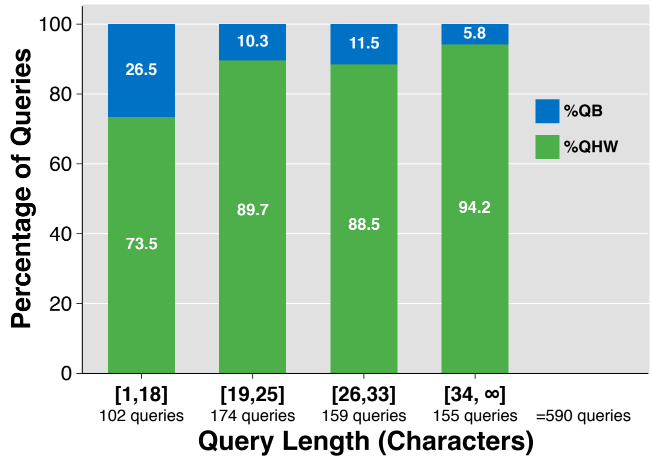

Figure 4: Overview of $\mathcal{QHW}$ vs. $\mathcal{QB}$ usage when reissuing queries of varying lengths. Shown here are the results over 590 reissued queries across all participants/conditions.

6.2 H2: Query Positioning in the QHW

H2 centres around the number of queries in $\mathcal{QHW}$. The hypothesis states that as the number of issued queries increases, the likelihood of a participant using $\mathcal{QHW}$ increases as the widget's distance to the starting point (❻ in Figure 3) decreases. We rank the five $\mathcal{QHW}$ positions (conditions) we explore empirically according to their distance from the point on the screen where we expect the participant's cursor to be after they have scanned all six results. This information is shown in row **II** of Table 3. The positions are ranked as follows: *(1)* `BOTTOM-LEFT`; *(2)* `MIDDLE-LEFT`; *(3)* `LOWER-RIGHT`; *(4)* `TOP-LEFT`; and *(5)* `UPPER-RIGHT`. From our interaction logs, we also calculated the position in the $\mathcal{QHW}$ for a reissued query. As discussed in §5.3, queries are displayed in reverse chronological order, which means if users have to scan further down the list, they are looking for an older query to reissue. We collected the ranks of all reissued queries (590 queries in total), and divided them into two groups based on how many queries are displayed *above the fold* in $\mathcal{QHW}$—reissued queries with a position of four or less (340 queries) and those with a rank greater than four (250 queries). Table 3, row **XI** shows that on average, participants are more likely to reissue a query when it is present *below the fold* from the conditions where $\mathcal{QHW}$ was positioned closer to the starting position (as observed in row **II**). Moreover, when participants want to reissue a recent

Table 3: Mean (± standard deviation) of search behaviour metrics across all participants in each variation of $\mathcal{QHW}$. A dagger (†) denotes one-way ANOVA significance, § denotes χ^2 significance, while $^{\mathcal{U}}, ^{\mathcal{L}}, ^{\mathcal{B}}, ^{\mathcal{M}}, ^{\mathcal{T}}$ indicate post-hoc significance ($p < 0.05$ with Bonferroni correction) over conditions UPPER-RIGHT, LOWER-RIGHT, BOTTOM-LEFT, MIDDLE-LEFT and TOP-LEFT respectively.

	Measure	UPPER-RIGHT	LOWER-RIGHT	BOTTOM-LEFT	MIDDLE-LEFT	TOP-LEFT
I	Number of participants	24	24	25	24	23
II	Rank in terms of distance from ❻ (D_{QHW}, pixels)	5(980)	3(610)	1(190)	2(515)	4(780)
Search Log Statistics						
III	Number of queries via $\mathcal{QB}$†	$10.29(\pm5.19)^{\mathcal{B}}$	$11.67(\pm5.04)$	$13.92(\pm6.05)^{\mathcal{U}}$	$12.29(\pm5.18)$	$12.61(\pm5.26)$
IV	Number of queries re-issued via $\mathcal{QHW}$	$3.50(\pm4.58)$	$3.79(\pm5.16)$	$5.28(\pm7.26)$	$5.21(\pm6.67)$	$3.78(\pm4.21)$
V	Number of queries re-issued via $\mathcal{QB}$	$0.62(\pm1.17)$	$0.67(\pm1.01)$	$0.60(\pm1.50)$	$0.54(\pm0.78)$	$0.65(\pm1.03)$
VI	Number of unique queries via $\mathcal{QB}$	$9.67(\pm4.72)^{\mathcal{B}}$	$11.00(\pm5.06)$	$13.32(\pm5.84)^{\mathcal{U}}$	$11.75(\pm4.79)$	$11.96(\pm5.43)$
VII	Number of hovers in $\mathcal{QHW}$ (500ms threshold)†	$5.50(\pm4.61)^{\mathcal{M},\mathcal{T}}$	$5.25(\pm6.07)^{\mathcal{T}}$	$6.72(\pm7.42)$	$9.50(\pm8.45)^{\mathcal{U}}$	$11.74(\pm6.09)^{\mathcal{U},\mathcal{L}}$
VIII	Number of scroll events on $\mathcal{QHW}$†	$2.75(\pm4.11)^{\mathcal{B},\mathcal{M}}$	$3.83(\pm6.15)$	$7.52(\pm10.03)^{\mathcal{U}}$	$7.42(\pm9.22)^{\mathcal{U}}$	$6.13(\pm7.82)$
IX	Frac. of queries re-issued via $\mathcal{QHW}$, slow typing (57 users)	$0.95(\pm0.12)$	$0.96(\pm0.07)$	$0.93(\pm0.12)$	$0.97(\pm0.03)$	$0.90(\pm0.18)$
X	Frac. of queries re-issued via $\mathcal{QHW}$, fast typing (57 users)	$0.78(\pm0.37)$	$0.88(\pm0.15)$	$0.90(\pm0.21)$	$0.80(\pm0.31)$	$0.75(\pm0.38)$
XI	Frac. queries re-issued *below* $\mathcal{QHW}$ fold§	$0.846^{\mathcal{B},\mathcal{M}}$	0.911	$0.979^{\mathcal{U}}$	$0.957^{\mathcal{U}}$	0.892
XII	Frac. queries re-issued *above* $\mathcal{QHW}$ fold	0.849	0.758	0.882	0.852	0.835
Post-Task Questionnaire						
XIII	Expected Position, *1: unexpected, 7: expected*†	$4.62(\pm1.24)^{\mathcal{M}}$	$4.22(\pm1.88)$	$4.64(\pm1.58)^{\mathcal{M}}$	$3.17(\pm2.12)^{\mathcal{U},\mathcal{B},\mathcal{T}}$	$4.96(\pm1.33)^{\mathcal{M}}$
XIV	Task Support, *1: obstructive, 7: supportive*	$5.25(\pm1.39)$	$5.30(\pm1.46)$	$5.40(\pm1.26)$	$4.33(\pm2.01)$	$4.91(\pm1.76)$
XV	Ease of use, *1: complicated, 7: easy*	$6.33(\pm1.20)$	$6.17(\pm1.23)$	$5.88(\pm0.93)$	$5.42(\pm1.61)$	$6.04(\pm1.30)$
XVI	Help in task goal, *1: inefficient, 7: efficient*†	$5.58(\pm1.50)^{\mathcal{M}}$	$5.04(\pm1.55)^{\mathcal{M}}$	$4.88(\pm1.59)$	$3.83(\pm2.33)^{\mathcal{U},\mathcal{L},\mathcal{T}}$	$5.48(\pm1.41)^{\mathcal{M}}$
XVII	Purpose of widget, *1: confusing, 7: clear*	$5.21(\pm2.28)$	$6.17(\pm1.50)$	$6.20(\pm1.32)$	$5.33(\pm1.74)$	$5.26(\pm1.91)$

query (displayed above the fold in $\mathcal{QHW}$), they are on average less likely to reissue it from the respective $\mathcal{QHW}$ conditions compared to when participants want to reissue an older query (displayed below the fold), as shown in Table 3, rows **XI** and **XII**. The only exception is the farthest variant UPPER-RIGHT, where the likelihood of reissuing is similar for both recent and older queries. To observe if this trend showing evidence for our hypothesis is significant, we conducted a Chi-squared test following a similar approach to **H1**. We sample two reissued queries (one above and one below the fold of $\mathcal{QHW}$) from each participant, and observe that for queries lower down the list (below the fold) in $\mathcal{QHW}$ (row **XI**), there is a significant difference for the fraction of time it was reissued via $\mathcal{QHW}$ across conditions ($\chi^2(4, N = 89) = 9.14, p = 0.02$). Post-hoc tests revealed significant differences between BOTTOM-LEFT and MIDDLE-LEFT in comparison to the UPPER-RIGHT condition. There was no significant difference when the query was present above the fold in $\mathcal{QHW}$ ($\chi^2(4, N = 114) = 3.86, p = 0.42$). Since we do not find a significant difference across all variants for queries reissued via $\mathcal{QHW}$ below the fold, we can only claim our hypothesis has been partially supported.

6.3 H3: Distance of the QHW

H3 states that with decreasing distance of the $\mathcal{QHW}$ to the starting point, the higher its usage. As mentioned in **H2**, row **II** of Table 3 shows the distance of each condition of $\mathcal{QHW}$ from where we expect a participant's mouse cursor to be after scanning results (❻ from Figure 3). We observe from row **IV** of Table 3 that there is no significant difference ($F(4, 115) = 0.544, p = 0.7$) in the number of reissued queries via $\mathcal{QHW}$ between the conditions. We do however observe a trend: participants in the two conditions closest to the starting point (BOTTOM-LEFT and MIDDLE-LEFT) issued on average more than five queries via $\mathcal{QHW}$; in the other three conditions, participants issued on average fewer than four queries via $\mathcal{QHW}$. Finally, it is worth mentioning that our $\mathcal{QHW}$ widget has the intended effect: reissued queries are far more likely to come via $\mathcal{QHW}$ than $\mathcal{QB}$ whose usage for reissued queries is shown in row **V** of Table 3. On average, fewer than one reissued query per participant is submitted via $\mathcal{QB}$. Based on these results we cannot argue in favour of **H3**, even though the data trend is aligned with our hypothesis, the differences are not significant.

6.4 H4: Slow Typing

H4 states that participants who type slower incur a higher typing cost, and are likely to prefer to use $\mathcal{QHW}$ irrespective of its position on screen. From our interaction logs, we computed the mean typing speed of our participants (we considered 114 users who reissued at least one previous query) by averaging the total time they took to type a query by the query length. Subsequently, we performed a median split (0.323 seconds per character) of our participants based on their mean typing speed, and categorised them as *Slow* and *Fast* at typing.

In rows **IX** and **X** of Table 3, we report the fraction of times reissued queries were issued via $\mathcal{QHW}$ by *Slow* and *Fast* participants, respectively. Across all conditions, we find, on average, *Slow* participants relied on $\mathcal{QHW}$ more often than those in *Fast*. For example, in the UPPER-RIGHT condition, 95% of reissued queries on average were submitted via $\mathcal{QHW}$ over *Slow*; the value was 78% for *Fast*. The smallest difference in behaviour is observed for the BOTTOM-LEFT condition: here, *Slow* reissue on average 93% of queries via $\mathcal{QHW}$, while *Fast* reissue 90% via $\mathcal{QHW}$. In addition, we find *Fast* to exhibit more diverse behaviour than *Slow* as indicated by the standard deviations reported in rows **IX** and **X** of Table 3. This shows that participants in *Slow* rely on $\mathcal{QHW}$ more consistently than *Fast*. There is no significant difference for the fraction of time a query was reissued via $\mathcal{QHW}$ by *Slow* (row **IX**). Although this finding is in line with our hypothesis, we do not see any significant difference over *Fast* (**X**). We thus cannot claim to have support for **H4**.

6.5 H5: F-Shaped Gaze Pattern

H5 is not derived from our formal model, but instead based on prior works that have found users to pay attention to SERPs in a particular manner: the top-left part of the SERP receives the most attention, and attention decreases as one goes down the SERP on the left rail, and to the right rail. We hypothesise here that we can find a similar attention pattern for the different positions of $\mathcal{QHW}$.

Contrasting to Dumais et al. [15] where gaze patterns were recorded via eye trackers, we did not perform webcam-based eye tracking and thus have to rely on other interaction logs to estimate attention. As found by Rodden et al. [45], eye gaze and mouse movements are correlated. We thus approximate how much attention participants were paying to $\mathcal{QHW}$ variants via the mean number of hover and scrolling events over $\mathcal{QHW}$ based on our interaction data[5]. We only consider hover events that spanned at least 500ms. We make this choice as variants `MIDDLE-LEFT` and `TOP-LEFT`, due to their location, would fall *'in the way'* of participants performing other tasks, like marking a document, or moving to $\mathcal{QB}$—considering all hover events would have skewed the interaction data. We observe significant differences across $\mathcal{QHW}$ variants for these two hover-based ($F(4, 115) = 41.4, p = 0.003$) and scroll events ($F(4, 115) = 39.6, p = 0.01$) which are reported in Table 3 (rows **VII** and **VIII**). Post-hoc tests revealed that the `UPPER-RIGHT` condition receives significantly fewer hovers or scrolls (and thus less attention) than `MIDDLE-LEFT`, `TOP-LEFT` and `BOTTOM-LEFT`. Thus, attention decreases as we move to the right. In contrast, we do not confirm our hypothesis that attention decreases as users move down the screen: the `TOP-LEFT` and `BOTTOM-LEFT` conditions do not significantly differ in terms of our hover/scroll measures.

Overall, we have partial evidence for hypothesis **H5**: participants pay less attention to the right side of the screen as approximated by our hover/scroll measures, but this attention decrease is not observed as we move towards the bottom of the screen. Of course, it should be noted that we designed the interface in such a way that participants were able to see the entire SERP at once without the need to scroll and see below the fold.

7 CONCLUSIONS

In this work, we set out to answer the question of *how to incorporate interface positioning information in a SET-based model.* To this end, we derived a position-aware interaction model of search behaviour. We focused on the *Query History Widget* ($\mathcal{QHW}$), and formulated a model that can predict search behaviour related to the *reissuing* of queries from the same search session. We used Fitts' Law [17] to approximate the cost of finding the widget based on its five different positions on the screen. Based on our model and prior works, we developed five testable hypotheses. We conducted a between-subjects user study with $N = 120$ participants. We evaluated the impact of the *position* of $\mathcal{QHW}$ on search behaviour.

[5]We are aware of more advanced approaches to estimate gaze patterns from mouse movements, e.g. [19]—but leave this exploration for future work.

Of our five hypotheses, we found partial support for three.

H1 As the length of the to-be-reissued query increases, a user will be more likely to reissue the query via the $\mathcal{QHW}$.
H2 If the number of queries to check in the $\mathcal{QHW}$ increases, the likelihood of users using the $\mathcal{QHW}$ increases as its distance to the starting point decreases.
H5 A user's attention span follows a F-shaped gaze pattern.

For the remaining two hypotheses—considering the relationship of the distance of $\mathcal{QHW}$ to the starting point and the widget's usage (**H3**), as well as the impact of typing speed on $\mathcal{QHW}$'s usage (**H4**)—we observed trends aligned with our hypotheses. However, those trends were not statistically significant. Our empirical study therefore did *not* provide support for them.

Overall we argue that we successfully developed a position-aware interaction model of search behaviour. We did find that widget positioning plays a role and changes a user's search behaviour, and thus *position matters*—and should be incorporated into formal interaction models. Our model is purposefully simple and does not capture every possible facet of user interaction with a SERP and its widgets. Several additions and modifications can be made.

Generalisation We focused on a simplified use case of a single widget (the $\mathcal{QHW}$). However, a modern search interface often contains multiple complex widgets simultaneously. Therefore, we aim to extend our work by creating user interaction models for more complex decisions pertaining to other widgets.
Cognitive effort Currently, our model ignores the cognitive cost of typing a query or looking for a query from a list present in the $\mathcal{QHW}$. Modelling cognitive costs is not trivial and depends on, amongst other factors, the search phase the user is currently in, a user's prior knowledge—and task difficulty.
Layout and Graphics We have assumed that *only* the location of the $\mathcal{QHW}$ and $\mathcal{QB}$ impact the cost of finding these widgets. However, it has been shown that during web navigation, there is a difference between ease of finding a graphical widget (e.g., a shopping basket in an e-commerce website or a search box) versus one that is textual (e.g., various text hyperlinks in navigation menus) [21, 23]. The graphical properties of these widgets, like size, shape, colour, and highlights, can also impact the efficiency in finding links and widgets [21, 24, 41, 47]. They likely provide certain *cues* [14] that users latch onto.
Usability and Aesthetics Based on our current model, a widget that takes 90% of the SERP would be straightforward for a user to find. However, it would also make the whole user experience unpleasant at best. Therefore, modelling user interactions with multiple widgets could help us strike a balance to optimise the complete user experience. Prior works focusing on developing aesthetic measures (i.e., based on colour and symmetry) for widgets can also help develop a more nuanced model [39, 54].
Input Devices Our model assumes that the user interacts with a search engine in a standard browser, using a mouse and keyboard. However, this is not always the case. Extending our model to other types of interfaces like mobile and voice search (building on existing work [28]) is another interesting research venue to explore in future work.

REFERENCES

[1] L. Azzopardi. 2011. The economics in interactive information retrieval. In *Proc. 34th ACM SIGIR*. 15–24.

[2] L. Azzopardi. 2014. Modelling interaction with economic models of search. In *Proc. 37th ACM SIGIR*. 3–12.

[3] L. Azzopardi, D. Kelly, and K. Brennan. 2013. How query cost affects search behavior. In *Proc. 36th ACM SIGIR*. 23–32.

[4] L. Azzopardi and G. Zuccon. 2015. An analysis of theories of search and search behavior. In *Proc. 1st ACM ICTIR*. 81–90.

[5] L. Azzopardi and G. Zuccon. 2016. An analysis of the cost and benefit of search interactions. In *Proc 2nd ACM ICTIR*. 59–68.

[6] L. Azzopardi and G. Zuccon. 2016. Two scrolls or one click: A cost model for browsing search results. In *Advances in Information Retrieval*. 696–702.

[7] L. Azzopardi and G. Zuccon. 2018. Economics models of interaction: a tutorial on modeling interaction using economics. (18 Jan. 2018).

[8] M. J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review* (1989).

[9] N. J. Belkin. 1990. The Cognitive Viewpoint in Information Science. *J. Inf. Sci.* 16, 1 (Jan. 1990), 11–15.

[10] Kuhlthau C. C. 1988. Developing a Model of the Library Search Process: Cognitive and Affective Aspects. *RQ* 28, 2 (1988), 232–242.

[11] S. K. Card, T. P. Moran, and A. Newell. 1980. The keystroke-level model for user performance time with interactive systems. *Commun. ACM* 23, 7 (1980), 396–410.

[12] B. Carterette. 2011. System effectiveness, user models, and user utility: a conceptual framework for investigation. In *Proc. 34th ACM SIGIR*. 903–912.

[13] Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. 2001. What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In *CHI'01 extended abstracts on Human factors in computing systems*. 281–282.

[14] E.H. Chi, P. Pirolli, K. Chen, and J. Pitkow. 2001. Using Info. Scent to Model User info. Needs and Actions and the Web. In *Proc. 19th ACM CHI*. 490–497.

[15] S. T. Dumais, G. Buscher, and E. Cutrell. 2010. Individual Differences in Gaze Patterns for Web Search. In *Proc. 3rd IIiX*. 185–194.

[16] D. Ellis. 1993. Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach. *The Library Quarterly: Information, Community, Policy* 63, 4 (1993), 469–486.

[17] P. M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *J. experimental psychology* 47, 6 (1954), 381.

[18] N. Fuhr. 2008. A probability ranking principle for interactive information retrieval. *Information Retrieval* 11, 3 (2008), 251–265.

[19] Q. Guo and E. Agichtein. 2010. Towards predicting web searcher gaze position from mouse movements. In *Proc. 28th ACM CHI*. 3601–3606.

[20] A. Hinderks, M. Schrepp, and J. Thomaschewski. 2018. A Benchmark for the Short Version of the User Experience Questionnaire.. In *WEBIST*. 373–377.

[21] G. Hinesley, M. H. Blackmon, and M. J. Carnot. 2008. The Importance of Graphics: Implications for Educational Hypertext Material. In *EdMedia+ Innovate Learning*. 1412–1421.

[22] P. Ingwersen and K. Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer.

[23] S. Karanam, H. van Oostendorp, M. C. P. Melguizo, and B. Indurkhya. 2013. Interaction of textual and graphical information in locating web page widgets. *Behaviour & Information Technology* 32, 5 (2013), 503–515.

[24] S. Karanam, J. Viswanathan, A. Theertha, B. Indurkhya, and H. Van Oostendorp. 2010. Impact of placing icons next to hyperlinks on information-retrieval tasks on the web. In *Proc. 32nd CogSci*, Vol. 32.

[25] N. Karousos, C. Katsanos, N. Tselios, and M. Xenos. 2013. Effortless tool-based evaluation of web form filling tasks using keystroke level model and fitts law. In *Proc. 31st ACM CHI*. 1851–1856.

[26] A. Kashyap, V. Hristidis, and M. Petropoulos. 2010. Facetor: cost-driven exploration of faceted query results. In *Proc. 19th ACM CIKM*. 719–728.

[27] D. Kieras et al. 2001. Using the keystroke-level model to estimate execution times. *University of Michigan* 555 (2001).

[28] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. 2017. What snippet size is needed in mobile web search?. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. 97–106.

[29] B. Laugwitz, T. Held, and M. Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symp. Austrian HCI & usability eng.* 63–76.

[30] C. J. Lin and S. Ho. 2020. Prediction of the use of mobile device interfaces in the progressive aging process with the model of Fitts' law. *J. Biomedical Informatics* 107 (2020), 103457.

[31] F. Loumakis, S. Stumpf, and D. Grayson. 2011. This Image Smells Good: Effects of Image Information Scent in Search Engine Results Pages. In *Proc. 20th ACM CIKM*. 475–484.

[32] I S. MacKenzie and W. Buxton. 1992. Extending Fitts' law to two-dimensional tasks. In *Proc. 10th ACM CHI*. 219–226.

[33] D. Maxwell and L. Azzopardi. 2014. Stuck in traffic: How temporal delays affect search behaviour. In *Proc. 5th IIiX*. 155–164.

[34] D. Maxwell and L. Azzopardi. 2018. Information scent, searching and stopping. In *Advances in Information Retrieval*. 210–222.

[35] D. Maxwell, L. Azzopardi, and Y. Moshfeghi. 2017. A Study of Snippet Length and Informativeness: Behaviour, Performance and User Experience. In *Proceedings of the 40th ACM SIGIR*. 135–144.

[36] D. Maxwell and C. Hauff. 2021. LogUI: Contemporary Logging Infrastructure for Web-Based Experiments. In *Advances in Information Retrieval*. 525–530.

[37] M. McGregor, L. Azzopardi, and M. Halvey. 2021. Untangling cost, effort, and load in information seeking and retrieval. In *Proceedings of the 2021 CHIIR*. 151–161.

[38] A. Moffat, F. Scholer, and P. Thomas. 2012. Models and metrics: IR evaluation as a user process. In *Proc. 17th ADCS*. 47–54.

[39] D. C. L. Ngo, L. S. Teo, and J. G. Byrne. 2003. Modelling interface aesthetics. *Information Sciences* 152 (2003), 25–46.

[40] K. Ong, K. Järvelin, M. Sanderson, and F. Scholer. 2018. Qwerty: The effects of typing on web search behavior. In *Proc. 3rd ACM CHIIR*. 281–284.

[41] R. Pearson and P. Van Schaik. 2003. The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task. *Intl. J. Human-Computer Studies* 59, 3 (2003), 327–353.

[42] P. Pirolli. 2007. *Information foraging theory: Adaptive interaction with information*. Oxford University Press.

[43] P. Pirolli and S.K. Card. 1999. Information Foraging. *Psychological Review* 106 (1999), 643–675.

[44] S.R. Putra, F. Moraes, and C. Hauff. 2018. SearchX: Empowering Collaborative Search Research. In *Proc. 41st ACM SIGIR*. 1265–1268.

[45] K. Rodden, X. Fu, A. Aula, and I. Spiro. 2008. Eye-mouse coordination patterns on web search results pages. In *Proc. 26th ACM CHI*. 2997–3002.

[46] C. Sutter and M. Ziefle. 2004. Psychomotor efficiency in users of notebook input devices: Confirmation and restrictions of Fitts' law as an evaluative tool for user-friendly design. In *Proc. 48th HFES*, Vol. 48. SAGE Publications Sage CA: Los Angeles, CA, 773–777.

[47] F. P. Tamborello and M. D. Byrne. 2005. Information search: the intersection of visual and semantic space. In *Proc. 23rd ACM CHI*. 1821–1824.

[48] H. R. Varian. 1999. Economics and Search. *SIGIR Forum* 33, 1 (Sept. 1999), 1–5.

[49] M. Verma and E. Yilmaz. 2017. Search Costs vs. User Satisfaction on Mobile. In *Advances in Information Retrieval*. 698–704.

[50] E.M. Voorhees. 2006. Overview of the TREC 2005 Robust Retrieval Track.. In *Proceedings of TREC-14*.

[51] T. Wilson. 2000. Human Information Behavior. *Informing Science* 3 (01 2000), 49–55.

[52] W. Wu, D. Kelly, and A. Sud. 2014. Using information scent and need for cognition to understand online search behavior. In *Proc. 37th ACM SIGIR*. 557–566.

[53] L. Xu, X. Zhou, and U. Gadiraju. 2020. How Does Team Composition Affect Knowledge Gain of Users in Collaborative Web Search?. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. 91–100.

[54] X. S. Zheng, I. Chakraborty, J. J.W. Lin, and R. Rauschenberger. 2009. Correlating low-level image statistics with users-rapid aesthetic and affective judgments of web pages. In *Proc 26th ACM CHI*. 1–10.

[55] G. K. Zipf. 1949. *Human Behaviour and the Principle of Least Effort*. Addison-Wesley.