Eye-Tracking Analysis of User Behavior and Performance in Web Search on Large and Small Screens

Jaewon Kim
Research School of Computer Science, The Australian National University, Building 108 (CSIT), Canberra, Australian Capital Territory 0200, Australia. E-mail: jaewon.kim@anu.edu.au

Paul Thomas
Research School of Computer Science, The Australian National University, Building 108 (CSIT), Canberra, Australian Capital Territory 0200, Australia and CSIRO, GPO Box 664, Canberra, Australian Capital Territory 2601, Australia. E-mail: paul.thomas@csiro.au

Ramesh Sankaranarayana and Tom Gedeon
Research School of Computer Science, The Australian National University, Building 108 (CSIT), Canberra, Australian Capital Territory 0200, Australia. E-mail: {ramesh.sankaranarayana, tom.gedeon}@anu.edu.au

Hwan-Jin Yoon
Statistical Consulting Unit, The Australian National University, John Dedman Building 27, Canberra, Australian Capital Territory 0200, Australia. E-mail: hwan-jin.yoon@anu.edu.au

In recent years, searching the web on mobile devices has become enormously popular. Because mobile devices have relatively small screens and show fewer search results, search behavior with mobile devices may be different from that with desktops or laptops. Therefore, examining these differences may suggest better, more efficient designs for mobile search engines. In this experiment, we use eye tracking to explore user behavior and performance. We analyze web searches with 2 task types on 2 differently sized screens: one for a desktop and the other for a mobile device. In addition, we examine the relationships between search performance and several search behaviors to allow further investigation of the differences engendered by the screens. We found that users have more difficulty extracting information from search results pages on the smaller screens, although they exhibit less eye movement as a result of an infrequent use of the scroll function. However, in terms of search performance, our findings suggest that there is no significant difference between the 2 screens in time spent on search results pages and the accuracy of finding answers. This suggests several possible ideas for the presentation design of search results pages on small devices.

Introduction

The enormous increase in the volume of information on the Internet has allowed users to access a range of data and retrieve appropriate information. This has given rise to the issue of developing search engines for user search performance, which can be defined in terms of search speed and search accuracy (Cutrell & Guan, 2007; Jones, Buchanan, & Thimbleby, 2003; Palmquist & Kim, 2000; van Schaik & Ling, 2001). Because search behavior is closely related to search performance (Buscher, Cutrell, & Morris, 2009; Granka, Joachims, & Gay, 2004), it is important to understand users' web search behavior in order to design effective and efficient search tools. (In this article, "behavior" includes a user's overall interaction with the search results page as well as eye fixations, reading and scanning patterns, clicks, and scrolls.)

Research into user behavior when conducting web searches, long studied in the field of human–computer

Received July 2, 2013; revised October 22, 2013; accepted October 23, 2013

© 2014 ASIS&T • Published online 11 June 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23187

interaction (HCI), has been performed by investigating interactions between the user and the web server (Buscher, White, Dumais, & Huang, 2012; Jansen & Spink, 2006; Silverstein, Henzinger, Marais, & Moricz, 1998) and by exploring users' search behavior through individual interviews or diary studies (Kelly, 2006; Teevan, Alvarado, Ackerman, & Karger, 2004). However, most recent studies have used eye-tracking technology to determine how users interact with each element of the web search results pages (e.g., Buscher et al., 2009; Cutrell & Guan, 2007; Granka et al., 2004; Lorigo et al., 2006). These studies have broadly concluded that there is some implicit meaning behind user search behavior, and this can be used to improve search engine performance.

Web searches on mobile devices have become common due to their convenience. People tend to access the Internet to search for information by using handheld devices such as smartphones or personal digital assistants (PDAs), even when laptops and desktops are available. However, current search engines do not provide different content for small devices, instead essentially simplifying their results pages.

Is there any difference between user search behavior on large and small screens? If there is, it may be worth redesigning the search results pages to support users' search performance by considering the differences in large and small devices. In this article, we explore user behavior and performance in web search on small and large screens, and analyze the relationships between search speed and several search behaviors in order to understand the difference in detail. Although several studies have investigated differences in user interactions within web search results between large and small screens for better design of search results pages (e.g., Findlater & McGrenere, 2008; Jones et al., 2003; Jones, Marsden, Mohd-Nasir, Boone, & Buchanan, 1999), the research results have not yet embraced search behavior captured by eye movements. Recently, eye-tracking technology has begun to be applied to eye movements relative to small devices (Biedert, Dengel, Buscher, & Vartan, 2012; Drewes, De Luca, & Schmidt, 2007; Nagamatsu, Yamamoto, & Sato, 2010). We focus on the influence of different screen sizes by using an emulator for the screen size of mobile devices on a desktop monitor. This is intended to prevent any recording interruption due to, for example, controlling a screen by touch. Our goal is to investigate whether there are differences in search behavior across screen sizes and, if so, to present ideas to design the presentation of search results to facilitate faster search with higher accuracy on small screens.

We first survey previous studies regarding search behavior on the web, and describe our experimental design and procedure. We then discuss how to measure this behavior. Our results and a discussion of the differences found with our screen sizes are addressed. We conclude by considering the implications of our findings, proposing some ideas for designing presentation of search results on handheld devices as future studies.

Literature Review

In this section, we introduce some of the background knowledge necessary for conducting this experiment. Four general lines of study should be considered: potential outcomes from an *eye-tracking study*, *general search behavior*, *search strategies* on web search results pages, and *user interaction on small screens*.

Eye-Tracking Studies

Eye tracking provides clues to user cognition as well as user interaction in various fields of computer science (Jacob & Karn, 2003; Rayner, 1998). In particular, for studies regarding web search, numerous papers have stated that eye tracking seems to facilitate our understanding of users' attention, because the gaze can show which elements of web search results pages receive attention (e.g., Aula, Majaranta, & Räihä, 2005; Buscher, Dumais, & Cutrell, 2010; Dumais, Buscher, & Cutrell, 2010; Cutrell & Guan, 2007; Granka et al., 2004). Therefore, eye tracking is a promising method for the experimental investigation of search behavior on mobile devices.

Eye tracking is too large a research area to cover comprehensively in this section. Instead, we look at the major factors relevant to this experiment. The main observations made when using eye tracking are *fixations* and *saccades*. Fixations can be defined as the moments when the eyes are relatively stationary in order to take in some information; fixation duration can be as short as 50–75 ms and as long as 500–600 ms. Saccades are the continuous eye movements between fixations. Velocities as high as 500 degrees per second have been observed. Fixations can have numerous meanings; saccades, even though it is widely believed that they do not say anything about a user's perception, may provide some clues to search behavior from their number or direction (for more details, see Rayner, 1998, 2009; Poole & Ball, 2006). We can define scanpaths that depict a complete sequence of both fixation and saccades.

The eye-movement behaviors just described have several implications for understanding search behavior. Goldberg and Kotval (1999) found that more fixations indicates less effective searching, and that the optimal scanpath in a search task exhibits a short fixation duration and less hesitation. In addition, more fixations on a particular area of interest implies that the information there is more important than that in other areas (Poole, Ball, & Phillips, 2005), whereas a longer average fixation duration is an indication of task complexity (Just & Carpenter, 1976; Rayner, 1998).

General Search Behaviors

Several approaches have been used to understand users' search behavior on the web. One early method involved

DOI: 10.1002/asi

analyzing transaction log files, which contain information about click-throughs, queries, and the scrolling interaction between users and search engines (Jansen & Spink, 2006; Silverstein et al., 1998). Silverstein et al. (1998) analyzed a large query log file of requests to determine how users interact with a commercial search engine. They suggested that searchers primarily scan the first 10 search results and rarely modify their query. Similarly, Jansen and Spink (2006) found that searchers view fewer results to obtain the information they need than in the past, and that longer time is spent on search result pages than on other web documents. They suggested that this may be due to a less complex interaction between users and search engines than in the past—a result of the general improvement in web search engines. More recently, Buscher et al. (2012) supplemented query logs with large-scale records of cursor movements and text highlighting from a commercial web search engine. They found that users who spent a short time on search results pages tended to inspect just a few results, scroll less, and use fast mouse movements.

Another approach uses diary studies and individual interviews to investigate user search behaviors, such as the impact of task complexity (Byström & Järvelin, 1995), orienteering behavior (Teevan et al., 2004), or context in online information seeking (Kelly, 2006).

In the study by Byström and Järvelin (1995), some useful results on general user behavior were highlighted, although web search behavior was not explicitly investigated. The links between task complexity, necessary information types, information channels, and sources were analyzed by classifying or categorizing their relationship. Data were collected by a combination of diaries and questionnaires. The authors found that the relationships among these factors were significantly systematic and logical. For example, as task complexity increased, users needed more information and a greater number of sources, whereas the success rate of finding the required information decreased.

Teevan et al. (2004) conducted an observational experiment to investigate the way in which users look for information on the web. This aimed to find an optimal search tool design via a modified diary study supplemented with direct observations and hour-long, semi-structured interviews. They described "orienteering," different from a "keyword search," as the search behavior whereby users obtain information using small steps without specific information. Their findings suggested that users often did not use keyword-based search engines as part of the orienteering strategy, and that orienteering behavior should be considered for web search tools.

Kelly (2006) investigated collection of data about information-seeking context, the aspects of this context, and relationships among these aspects in "natural online environments." Using a diary study approach, she observed the behaviors of seven subjects over a 14-week period. The results suggested that the task and topic had significant effects on the perception of usefulness of documents in helping users complete tasks.

Analyzing transaction log files can determine user interaction in web searches in terms of explicit actions (e.g., mouse clicks, queries or cursor movement) with rich statistical results, and diary reports are useful in information interaction studies. However, they cannot provide detailed information about where users are looking. This seems to be a limitation when attempting to determine why users interact with different elements of web pages moment by moment. If we can gain access to the gaze data, it is possible to not only analyze their search scanpath and scanning strategy, but to also compare the search behavior to explicit data such as search speed and click patterns. Thus, we must consider studies that have focused on eye tracking.

A great deal of research has been performed using eye tracking to investigate behavior on search results pages. Several studies attempted to find broad scanning patterns, such as a "golden triangle" for an optimal search engine design (Hotchkiss, Alston, & Edwards, 2005), an "F-shaped pattern" for web usability (Nielsen, 2006), or better web page design (Buscher et al., 2009), by measuring which elements of web pages caught the searchers' eye and the sequence of eye movements while searching.

Broder (2002) developed a taxonomy to investigate queries for the classification of user goals in web searches. This taxonomy of web searches was divided into three classes: *informational*, for finding information on one or two web pages; *navigational*, for reaching a particular page; and *transactional*, for performing a certain transaction. As a result, he suggested that current search engines need to determine the user's goal in order to maximize their satisfaction. Broder's classification has been broadly adopted.

Lorigo et al. (2006) studied differences in search behavior by gender and task type (informational and navigational tasks) using fixations and scanpaths. With scanpaths, they defined a compressed sequence and a minimal scanpath, and introduced three additional terms—complete, linear, and strictly linear (a detailed description of these terms is given in the next section). They found that the task type had no effect on the scanpaths or search accuracy, but suggested that the task type may impact the task completion time: users tend to spend more time finishing tasks when conducting informational, rather than navigational, tasks. They also suggested that users do not normally follow the rank order given by a search engine, with over 50% of subjects engaging in *regression* (jumping back at least one link) and *skipping* (jumping over one link) in a gaze sequence.

Some studies have focused on how searchers' scans are related to their click decision on search results pages. Joachims, Granka, Pan, Hembrooke, and Gay (2005) and Granka et al. (2004) have examined how eye fixations relate to scrolling, and how users browse search results above and below their final selections. Their results indicate that subjects rarely scan below the selected link except when the link is at the page fold (in which case users often scan further). They also found that users' fixations are significantly

DOI: 10.1002/asi

clustered within the results ranked one and two places before their selection. These findings suggested that users tend to move from top to bottom when scanning the result listings. Furthermore, Joachims et al. (2005) found that participants were influenced by the relevance of the results, indicating that users scanned more links, and clicked relatively lower ranks, in a "reversed" condition (with a poorer ranking) than in the "normal" condition. Guan and Cutrell (2007) suggested a similar result. They examined users' search behavior when the rank orders of relevant results were manipulated, and found that the search efficiency was greatly decreased when the relevant links were located in lower positions on the results pages. This is because the rank order affects search time and the rate of finding correct answers. The findings of the two studies were explained by the strong "trust bias" users have regarding the rank order of search results.

A study by Cutrell and Guan (2007) focused on the composition of three elements on search result pages (title, snippets, and URLs). They manipulated the elements in the presentation of the results pages to investigate the effects. As a consequence, they found that adding information to the snippet significantly enhanced search speed and accuracy for informational tasks, whereas it worsened performance for navigational tasks. They also observed a higher success rate in navigational searches.

Search Strategy

Several studies have classified users' search behavior according to gaze patterns. Klöckner, Wirschum, and Jameson (2004) found that 52–65% of participants used what they call a "depth-first" strategy (the subjects scanned only the links above the selected link, that is, they clicked it as soon as they saw an attractive link), 11–15% used a "breadth-first" strategy (the subjects looked through all the links before making a decision and selecting a link), and the remaining 20–37% showed a "mixed" strategy (looking ahead a few results past the link they selected). Aula et al. (2005) defined two kinds of search strategy in terms of users' evaluation patterns. They suggested that the 54% of subjects who scanned less than half of the visible results were "economic" evaluators and that the others had an "exhaustive" evaluation style.

Dumais et al. (2010) extended this classification and defined three clusters: "economic-results" users who look at few additional results below the link they selected, "economic-ads" users who regularly look at advertisements, and "exhaustive" users who scan the results broadly. Measures of fixation and scanpaths were used to identify the users' search strategies when viewing the results of major commercial search engines that include additional sponsored links or advertisements. According to their findings, the economic-results and economic-ads groups tended to spend more time on the first three results than did the exhaustive users (68%, 61%, and 53%, respectively), and the total fixation time of each group showed that the exhaustive participants reviewed the results most slowly. In addition, the exhaustive users had different scanpaths than the other two groups.

User Interaction on Small Screens

A few researchers have investigated explicit user interaction with small screens on web search results pages, although these studies did not record eye movements. For example, Jones et al. (2003) compared users' abilities among three kinds of interfaces: mobile phone-sized, handheld computer-sized (PDA), and conventional desktop. They found that users take more time to complete tasks and exhibit lower task success rates on smaller screens. They suggested several guidelines for the design of small screens to improve user search performance such as sufficient information in a search result, some marker to indicate that a link will display a small-screen-optimized page, and preprocessing conventional web documents for small devices.

Despite the interest in web search behavior, only a few eye-tracking studies have been conducted on small screens. Drewes et al. (2007) investigated gaze interaction for controlling applications on a handheld device using dwell time and gaze gestures. Further, Nagamatsu et al. (2010) investigated a remote gaze tracker for mobile devices with stereo gaze tracking. Recently, text interaction and reading performed on an actual mobile touch screen device was analyzed by Biedert et al. (2012). However, we could find no investigation that used eye tracking to study user behavior in web searches on small devices.

The research presented in this article extends previous studies with the aim of understanding the differences in user search performance and behavior on the web with respect to two differently sized screens and two task types—informational and navigational. As well as considering differences in behavior, we compare relationships between search performance and behavior across the two sizes of screens. Furthermore, we suggest some improvements that may enable the design of a better presentation method for search results pages on small screens.

Experimental Design

In this section, we explain the experimental design and our procedure. Using an eye-tracking instrument, mouse movement logging, and other instruments, we recorded gaze and click data from 32 participants as they interacted with web search results. Subjects completed a total of 640 search tasks, on both large and small screens, with the Google mobile search engine. The "large" screen showed 10 search results, whereas the "small" screen showed only about three results and was designed to have a screen size similiar to popular mobile devices.

Participants

Thirty-five subjects (19 male) between 18 and 50 years old, from various disciplines and recruited on campus at a

TABLE 1. Examples of task descriptions and queries.

Task description	Initial task query	Task type
Find the official homepage of the Canberra casino and hotel in Canberra.	Canberra Casino	Nav
Go to the homepage of the Canberra Cavalry baseball team.	Canberra cavalry baseball	Nav
What is the standard length of a cue used for playing billiards?	billiard cue size	Info
How many spikes are in the crown of the Statue of Liberty?	statue of liberty crown spikes	Info

Note. Nav denotes navigational task and Info denotes informational task.

local university, participated in the eye-tracking study. All subjects claimed to be experienced in web searching, and were quite familiar with the Google search engine. We excluded the results from three participants (2 male, 1 female) for technical reasons (e.g., stability problems with eye tracking).

Tasks

Each participant completed 20 search tasks. Following Lorigo et al. (2006), we adopted two task types—informational and navigational—each represented by 10 tasks to investigate the influence of the task type. Each task was shown to the participants with a task description and a predefined query. The descriptions and queries were obtained from Dumais et al. (2010) and then modified for local participants (see Table 1; see also Appendix A1 for further details). Search results were retrieved from the Google mobile search engine, from which we removed images, maps, and related links so that all tasks showed the same elements (see Figure 1). All results pages were cached as local files in the system and relevant pages were shown when subjects chose the links. The tasks were very simple, needing only 1–2 min to complete. The initial queries were effective: 18 of the predefined results pages contained a relevant solution within the top three results, with the other two including a relevant result in ranks 4–6.

Design

The participants were divided into four groups of eight, and the tasks were arranged in two sets: set 1 consisted of informational tasks 1–5 and navigational tasks 1–5 in rotation; set 2 contained the remaining tasks analogously (i.e., set 1 consisted of ⟨I1, N1, I2, N2, . . . N5⟩ and set 2 of ⟨I6, N6, I7, N7, . . . N10⟩, where “I” denotes an informational and “N” a navigational task). Each subject performed both task sets, one on a large screen and the other on a small screen, and both the set order and screen order were counterbalanced across subjects. In other words, subject 1 performed task set 1 (TS1) on the large screen and then task set 2 (TS2) on the small screen, followed by subject 2 performing TS2 on the large screen and then TS1 on the small screen, and so on (see Table 2). Therefore, each task was distributed 32 times (16 times on each screen size) across the participants.

Procedure

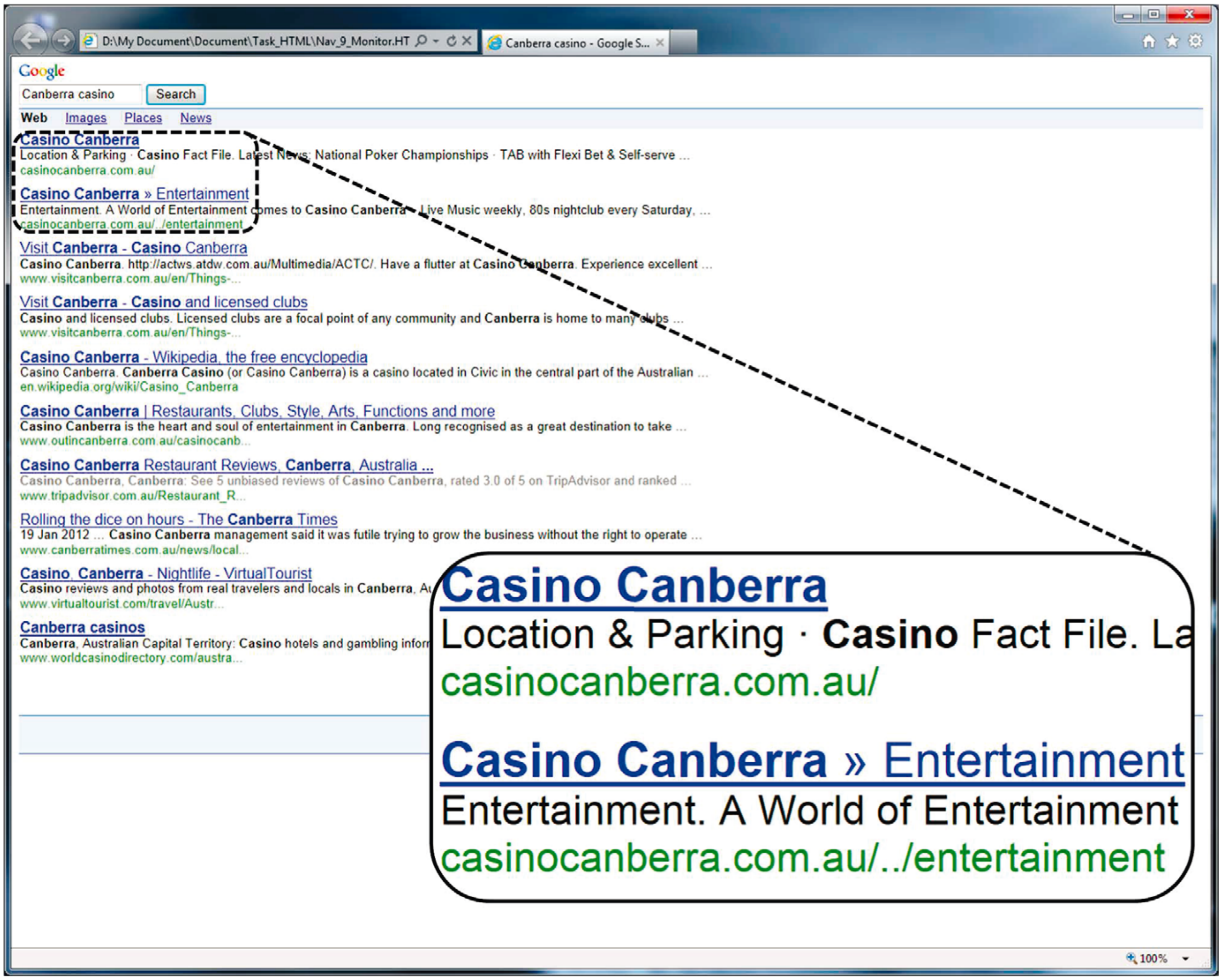
All participants first listened to an introduction to the experiment, and practiced two sample tasks on each screen until they were familiar with the system. Their head was then fixed on a chinrest to ensure higher eye-gaze detection accuracy, and the eye tracker was calibrated using 5-point calibration (see Figure 2). Next, we gave the participant the first task description, an initial query, and then showed the results page. This procedure was repeated for all 20 tasks according to an automated schedule. A time notice was given 3 min after starting each task, after which the subjects were free to either take more time to find the answer or move on to the next task. Participants were not allowed to type queries, as looking at the keyboard often caused eye tracking to be lost. However, as explained, the cached search results pages included sufficient relevant links to acquire the answers, and participants should not have required the keyboard during the experiment. Furthermore, they could continue to the next page of results or follow links from the list of results. The participants could ask for a full task description if they did not understand the task sufficiently. At the end of the experiment, the subjects were asked to complete a questionnaire about their web search experience. The experimental run time was approximately 30 min for each participant.

Apparatus

All search results were obtained from the Google mobile search engine and displayed in Internet Explorer 8. Eye gaze was recorded by Facelab 5 with a desk-mounted 17″ LCD monitor and a chinrest, and data analyses were performed using the Eyeworks software (http://www.seeingmachines.com/product–/facelab/).

The large screen ran at a default resolution of 1280 × 1024 pixels. To simulate the small screen of a mobile device, we used the same monitor but with a browser limited to a 320 × 480 pixel window. Although the experiment was not performed on an actual mobile device, the small browser was adjusted to show a font size and number of search results that were similar to a current smartphone. To compare the effects of different screen sizes, we adopted the same font on both browsers. With these settings, in the case of the large screen, there was no fold and the 10 search results were clearly visible without needing any user scrolling (see Figure 1), whereas on the small screen the fold normally occurred a little after the third search result (see Figure 3).

DOI: 10.1002/asi


D:\My Document\Document\Task_HTML\Nav_9_Monitor.HT
Canberra casino - Google S...
Google
Canberra casino
Search
Web
Images
Places
News
100%

Casino Canberra
Location & Parking · Casino Fact File. Latest News: National Poker Championships · TAB with Flexi Bet & Self-serve ...
casinocanberra.com.au/

Casino Canberra » Entertainment
Entertainment. A World of Entertainment comes to Casino Canberra - Live Music weekly, 80s nightclub every Saturday, ...
casinocanberra.com.au/../entertainment

Visit Canberra - Casino Canberra
Casino Canberra. http://actws.atdw.com.au/Multimedia/ACTC/. Have a flutter at Casino Canberra. Experience excellent ...
www.visitcanberra.com.au/en/Things-...

Visit Canberra - Casino and licensed clubs
Casino and licensed clubs. Licensed clubs are a focal point of any community and Canberra is home to many clubs ...
www.visitcanberra.com.au/en/Things-...

Casino Canberra - Wikipedia, the free encyclopedia
Casino Canberra. Canberra Casino (or Casino Canberra) is a casino located in Civic in the central part of the Australian ...
en.wikipedia.org/wiki/Casino_Canberra

Casino Canberra | Restaurants, Clubs, Style, Arts, Functions and more
Casino Canberra is the heart and soul of entertainment in Canberra. Long recognised as a great destination to take ...
www.outincanberra.com.au/casinocanb...

Casino Canberra Restaurant Reviews, Canberra, Australia ...
Casino Canberra, Canberra: See 5 unbiased reviews of Casino Canberra, rated 3.0 of 5 on TripAdvisor and ranked ...
www.tripadvisor.com.au/Restaurant_R...

Rolling the dice on hours - The Canberra Times
19 Jan 2012 ... Casino Canberra management said it was futile trying to grow the business without the right to operate ...
www.canberratimes.com.au/news/local...

Casino, Canberra - Nightlife - VirtualTourist
Casino reviews and photos from real travelers and locals in Canberra, Au
www.virtualtourist.com/travel/Austr...

Canberra casinos
Canberra, Australian Capital Territory: Casino hotels and gambling infor
www.worldcasinodirectory.com/austra...

Casino Canberra
Location & Parking · Casino Fact File. La
casinocanberra.com.au/

Casino Canberra » Entertainment
Entertainment. A World of Entertainment
casinocanberra.com.au/../entertainment

FIG. 1. Search results as shown on the large screen. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE 2. Examples of experiment design for each group.

	Task set, order, and screen size
Group 1 ($N = 8$)	TS1 on L, and then TS2 on S
Group 2 ($N = 8$)	TS2 on L, and then TS1 on S
Group 3 ($N = 8$)	TS1 on S, and then TS2 on L
Group 4 ($N = 8$)	TS2 on S, and then TS1 on L

Note. L denotes a large screen and S denotes a small screen.

Measurements

We took several measurements for each participant, task, and screen size, focusing on search performance and behavior.

Search Performance

As described in the Introduction, we considered participants' search performance via two factors: search speed,

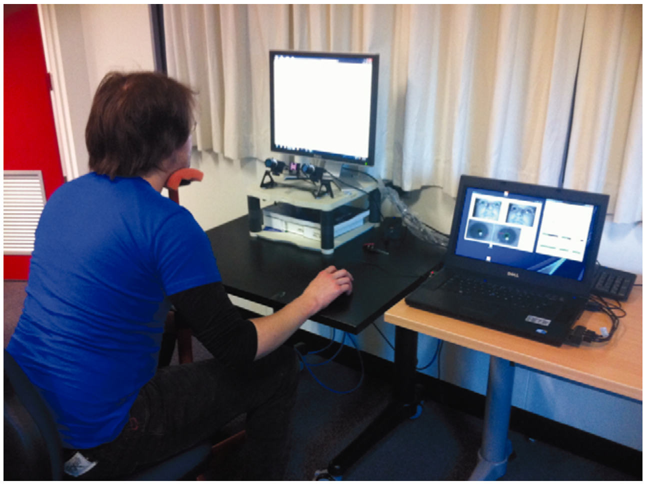
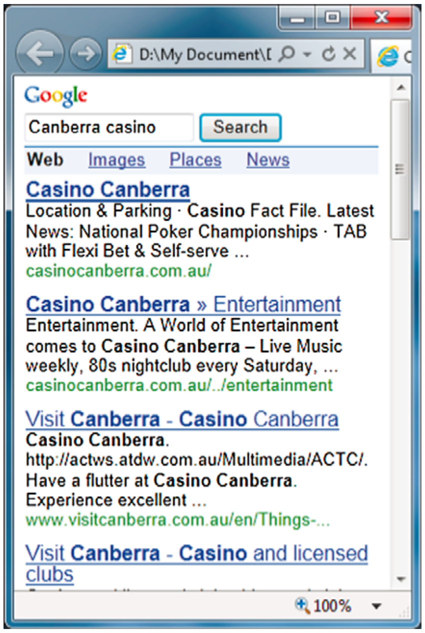


FIG. 2. A sample of the experimental environment. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



Canberra casino

Casino Canberra
Location & Parking · Casino Fact File. Latest News: National Poker Championships · TAB with Flexi Bet & Self-serve ...
casinocanberra.com.au/

Casino Canberra » Entertainment
Entertainment. A World of Entertainment comes to Casino Canberra – Live Music weekly, 80s nightclub every Saturday, ...
casinocanberra.com.au/../entertainment

Visit Canberra - Casino Canberra
Casino Canberra. http://actws.atdw.com.au/Multimedia/ACTC/. Have a flutter at Casino Canberra. Experience excellent ...
www.visitcanberra.com.au/en/Things-...

Visit Canberra - Casino and licensed clubs

FIG. 3. Search results as displayed on the small screen. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

measured mainly as time spent before the first click on the search results page and also as the task completion duration; and search accuracy, measured as the proportion of correct answers at the first attempt.

Search speed. Two explicit measurements of user search speed were recorded during the experiment: the elapsed time to the user's first click and the total duration of each task.

Because users' paths might diverge after their first click, and there was a 3-min time limit that was rarely reached before task completion, we calculated the gaze time spent prior to the first click as the major indicator of search speed. This differs from the approach of Cutrell and Guan (2007) and Jones et al. (2003), who measured the task completion duration to indicate the search speed. The gaze time can be calculated from when the contents are shown on the browser to the first click on a link. This indicates how long users spend making their decision.

The other measure, task completion duration, is the time between starting on a search results page and either finding a correct answer or reaching the time limit. This was recorded as a supplementary measurement of search speed.

Search accuracy. We considered the search accuracy to be 1 if users chose a relevant link that contained a correct answer on a search results page, and then found the answer on the first selected page. Otherwise, the search accuracy was assigned a score of 0. In the experiment, the subjects were informed that once they obtained the correct answer, they needed to provide this answer to the experimenter. If they chose not to do so, even after finding the correct answer on the first selected page, then the search accuracy would be (incorrectly) assigned a score of 0. We also reviewed another measurement of accuracy (following Cutrell & Guan, 2007) that counts how often users click "only one best result." However, determining the best link was impossible for informational tasks in our experiment, because there were several equally good candidates. (There is, of course, only one best result for the navigational tasks.)

Search Behavior

To study search behavior, we measured fixations and other gaze behavior, click patterns, scanpaths, scrolls, and similar interactions.

Fixation duration. Fixation duration on an area of interest (AOI) is a useful representation of how long a user spends obtaining information from a particular place (Dumais et al., 2010). As all initial search results pages had the same components (10 ranks and the periphery, e.g., a query box, category tabs, or blanks between AOIs), we assigned 10 AOIs to each search results page to investigate users' attention. Each AOI corresponded to a search result; that is, a clickable link along with its snippet text and a URL. The fixations were recorded if a gaze lasted at least 75 ms and if the gaze locations were close to each other (within a radius of five pixels), using the algorithms in the Eyeworks software. Although these values for the fixation duration and radius are relatively small compared to previous work (normally 100–300 ms and over 10 pixels; see, e.g., Granka et al., 2004; Rayner, 1998), we have optimized the figures for the font size of the initial search results pages presented by the Google mobile search engine, and the parameters are in line with those of Rayner (1998). We measured the mean fixation duration for each task and AOI, bearing in mind that a longer fixation duration indicates that it is more difficult to extract information from that task or AOI (Just & Carpenter, 1976; Rayner, 1998).

Click pattern. Click points were also recorded to analyze whether subjects chose correct answers, as well as how much bias the participants displayed with regard to result rank orders on each screen. In particular, users' first clicks were recorded as one method of determining search accuracy (see the subsection on search accuracy). We divided the click pattern into the top links (rank 1–3) and the others to investigate the bias and the frequency of scrollbar use.

Scanpath. Even though AOIs are invaluable in understanding where users are looking, and for how long, they do not by themselves provide any sequence information. Scanpaths capture this: A *raw* scanpath is simply a series of fixations on a results page, ordered by time. We also adopted the *compressed* and *minimal* scanpaths introduced by Lorigo et al. (2006). If we assume that the original scanpath is 2-2-1-1-2-3-3-4-5-5-4 (the numbers refer to the AOI rank of each fixation), the compressed sequence given by aggregating subsequent fixations is 2-1-2-3-4-5-4 and the length value is seven. That is, the compressed sequence describes how many ranks the user has viewed, including repeat visits. The minimal scanpath, given by removing repeat visits, is 2-1-3-4-5, which has a length value of five. In other words, the minimal path can be interpreted as how many different links a user has viewed before making a selection, as well as showing the overall sequence of fixations.

Scanning direction. With these two definitions, we may analyze three further types of scan pattern by refining the measurement methods of previous studies. Similar to previous work (Dumais et al., 2010; Lorigo et al., 2006), the first method describes a pattern as *complete* if the user inspected all of the links before making a selection. Next, a scanpath is said to be *linear* if the minimal path is monotonically increasing. This means that the user can only look ahead one or more ranks at a time, or back at ranks they have already observed in the original scanpath. Finally, a scanpath is *strictly linear* if the compressed sequence is monotonically increasing without any skips or regressions. In our measurements, a *skip* is defined as a jump of more than one rank (e.g., from rank 2 to 4) and a *regression* is a jump back of at least one rank (e.g., from 4 to 3).

In addition, because previous studies did not consider jumps of more than one link (e.g., from rank 3 to 5) in the definition of the linear pattern (or for not strictly linear), we added the condition that skips are allowed from the monotonically increasing point of view. For example, if we assume that the original scanpath is 1-1-2-4-2-5, the minimal scanpath is 1-2-4-5 and it increases monotonically. Thus it can be considered as a linear pattern, although it includes skips (2 to 4 and 2 to 5) and a regression (4 to 2) to a link which is already observed. However, this cannot be a strictly linear pattern, because the compressed sequence (1-2-4-2-5) does not increase monotonically but includes skips and a regression. Any pattern which is strictly linear is also linear, by definition.

Furthermore, past studies do not seem to include the case where users look at only one link and select it immediately. If users often make such a decision during the experiment, we should consider such cases. Therefore, we measured two linear and strictly linear patterns, which are named *linear* and *linear or/ID* (or *immediate decision*) and *strictly linear* and *strictly linear or/ID*. Any pattern that is linear or/ID is also strictly linear or/ID (and vice versa), and the compressed sequence will have length 1. For example, if a user only looks at rank 2, and then selects the link, the behavior is considered neither linear nor strictly linear, but is both linear or/ID and strictly linear or/ID.

Skip and regression. The compressed and minimal scanpaths also present skip and regression patterns. We decided to investigate skips and regressions in the data, as only the overall rate was given in a previous study (Lorigo et al., 2006). In this experiment, the proportion of skips and regressions and the distance between links were measured for each screen size and task type. We also analyzed cases in which users either looked at rank 1 or 2 before the skip (SAT: Skip After looking at Top ranks) or jumped back to rank 1 or 2 after reading further (RTT: Regression To Top ranks). This may give some insight into the users' trust in the ranking presented by the search engine.

Change of scan direction: ScanUp and ScanDown. As introduced by Dumais et al. (2010), changes in scan direction can be measured using *ScanUp* and *ScanDown* metrics, which record how often a user scanned up or down the ranked results until making their first selection. The scan downward sequence is determined if more than two subsequent compressed paths show the same downward direction (e.g., from position 2 to 3, and then from 3 to 4) or if there is a skip downward between two entries of a compressed path. The scan upward sequences are measured similarly. We counted the number of ScanDown and ScanUp occurrences to show the frequency of changes in scan direction. For example, if the compressed scanpath is 2-3-4-3-4-2-4, it has two ScanDown events and one ScanUp.

Maximum gaze position and scroll. Following Dumais et al. (2010), we recorded the maximum gaze position to gauge the highest rank looked at by users. This value is not only helpful in further analyzing the results of search strategy with trackback values (see subsection, Search Strategy), but also allows us to determine whether the scroll function is used during each small screen task.

As we can see from Figures 1 and 3, scrollbars only appear on the small screen. We measured the frequency of scrollbar use, which may indicate the usability of small screens, and investigated which links were selected after using the scroll bar, which may indicate the reliability of the search engine's ranking.

Search Strategy

A user's search strategy can be considered as an aggregate of their search behaviors. The classification of Aula et al. (2005) divides sessions into *economic* and *exhaustive* patterns depending on whether users scan at most half of the visible results without scrolling. This is not suitable for small devices with few visible links, as on a small screen, and instead we adopt the classification of Klöckner et al. (2004). Klöckner et al. described *depth-first*, *breadth-first*, and *mixed* patterns: In the depth-first pattern, users follow a promising link immediately; in the breadth-first pattern users study all options exhaustively before clicking; and in the mixed pattern, users read ahead, but to a smaller extent.

We used *trackback* (Kim, Thomas, Sankaranarayana, & Gedeon, 2012) to further investigate user search strategies in terms of how much a user reads before making a selection on the search results pages. Kim et al. (2012) defined this measurement to look into an aspect of the mixed strategy that had a similar frequency on both sized screens. Because most of our users exhibited a top-to-bottom scanning sequence (as

DOI: 10.1002/asi

in previous studies by Granka et al., 2004 and Joachims et al., 2005), we measured the distance between the selected link and the farthest link. For example, if a subject looked as far as AOI 7 and then clicked AOI 3, we recorded a trackback value of 4. In the rare cases of a user scanning from bottom to top, the trackback value is considered to be zero. This method has a little similarity to the maximum gaze point method and some relation to the scanpath analysis method. However, trackback is unique in that it summarizes the amount of additional effort users make before selecting a link.

Questionnaire Measures

Questionnaires are a useful supplement to interaction data, as they can elicit users' impressions and intentions during the experiment. After the experiment, subjects were asked several questions: gender, age, search convenience on each screen, level of difficulty of the tasks, self-reported search strategy on each screen, personal usage of Internet, satisfaction with the provided search engine, and favorite search engine. Some of this questionnaire data was compared with the recorded data: for example, to compare users' self-reported strategy to their actual performance.

Results and Discussion

Our data set consists of gaze data from 640 queries (320 queries on the large screen and 320 on the small screen, 160 informational and navigational queries on each sized screen). The analyses are focused on search performance (search speed and accuracy) and search behavior (mean fixation duration, scanpaths, and search strategies up to the first click). We consider two main effects, screen size and task type, and the interaction between these main effects. Using GenStat version 15 to analyze our data (VSN International, 2012), we employed analysis of variance (ANOVA) for continuous data such as times and durations with log-transformation if necessary, generalized linear models (GLMs) (McCullagh & Nelder, 1989) with binomial distribution and logit link function for binary data, and GLMs with Poisson distribution and logarithm link function for count data. We discuss the meaning of search speed and behavior, and also examine the relationships between search speed and some search behaviors to investigate the differences due to screen size.

Search Performance

The search performance results were analyzed in terms of search speed and search accuracy. The main result for search speed suggests that there is no significant screen size effect on the time taken until the first click on the results pages, although the supplementary results (task completion durations) for each screen are significantly different due to reading web documents after search results pages. In addition, the search accuracy results also show no difference across the screen sizes.

Search speed. The time elapsed until first clicks and task completion duration are shown in Table 3. One of the assumptions for ANOVA is normality. Because the data for time elapsed to first click and task completion duration do not meet the normality assumption, we used a log-transformation, $\log(x + 1)$, so that 0 maps to 0. We then employed ANOVA to investigate the main effects of screen size and task type, and the interaction between these effects. There is a significant task type effect ($F_{(1,605)} = 26.10$, $p < 0.001$) on the time taken until first click on the search results pages. Similar to the findings of Jones et al. (2003) on the effect of screen size and Lorigo et al. (2006) on the effect of task type, the task completion duration results show significant effects due to screen size ($F_{(1,605)} = 24.87$, $p < 0.001$), and task type ($F_{(1,605)} = 97.81$, $p < 0.001$), as well as an interaction effect ($F_{(1,605)} = 8.10$, $p < 0.01$). Both measurements of search speed suggest that subjects spent more time on informational tasks. In terms of screen size, the results indicate that users needed a longer time to complete tasks on the small screen, whereas there was no significant difference in time taken to make the first decision. The interaction of screen size and task type on task completion duration implies that users had more difficulty completing informational tasks on the small screen, because the difference by screen size was over 15 s (42.63 vs. 61.33 s), whereas there was little difference between completion duration for navigational tasks. We also calculated the number of page visits, finding no significant effect with respect to screen size but a noticeable effect depending on task type (see Table 3). The relation between task completion duration and page visits indicates that users tended to take extra time studying web documents for informational tasks on the small screen, because the page visit count is not significantly different on this screen size. This may be caused by less visible content, making it difficult to find particular information on the small screen, whereas subjects only needed to reach the correct pages for navigational tasks.

Search accuracy. For the accuracy data, we used a GLM with binomial distribution. The results show that there is only a significant effect with respect to task type ($\chi^2 = 17.81$, $df = 1$, $p < 0.001$). Similar to the findings of Cutrell and Guan (2007) on the effect of task type for search accuracy, the results in Table 3 show that users attained higher search accuracy for navigational tasks. This is because it was relatively difficult to reach the right answer for the informational tasks with the contents on the search results pages, whereas it was much easier to find the correct results for the navigational tasks from URLs or titles on either size of screen.

Subjects took a similar time until their first click and exhibited similar accuracy on both screen sizes despite fewer results being displayed on the small screen. However, the time spent on web documents differs significantly according to task type and screen size.

DOI: 10.1002/asi

TABLE 3. Search performance and behavior.

		Large		Small			*p* value	
		Info	Nav	Info	Nav	Screen size	Task type	Interaction
Search performance								
Search speed	Time to first click, s	18.9	15.9	20.65	16.41	$p = 0.342$	***	$p = 0.959$
	Task completion duration, s	42.63	29.85	61.33	30.64	***	***	*
Search accuracy	Correct click rate, %	69.38	78.75	62.50	82.50	$p = 0.655$	***	$p = 0.138$
Page visits	Counts (including SRPs)	2.84	2.49	2.84	2.29	$p = 0.420$	***	$p = 0.406$
Search behavior								
Fixation duration on SRP	Per task, s	3.06	2.60	4.01	3.21	***	***	$p = 0.138$
	Per link, s	0.91	0.86	1.27	1.12	***	*	$p = 0.291$
Clicks	On link 1–3 (%)	83.13	90.63	88.13	91.25	$p = 0.762$	***	$p = 0.509$
Scanpath	Minimal scanpath	3.46	3.11	3.13	2.92	$p = 0.061$	$p = 0.050$	$p = 0.068$
	Compressed sequence	5.22	4.82	4.78	4.64	$p = 0.073$	$p = 0.123$	$p = 0.481$
	Compressed –Minimal	1.76	1.70	1.65	1.72	$p = 0.629$	$p = 0.952$	$p = 0.545$
Scanning direction	Complete rate (%)	87.50	91.87	96.88	98.12	***	$p = 0.140$	$p = 0.958$
	Linear rate (%)	43.75	41.25	73.75	54.38	***	**	*
	Linear or/ID rate, %	56.25	63.75	90.00	85.63	***	$p = 0.635$	$p = 0.079$
	Strictly linear rate, %	13.13	14.38	33.75	15.63	***	**	**
	Strictly linear or/ID, %	25.63	35.00	49.36	45.00	***	$p = 0.510$	$p = 0.061$
Skip and regression	Skip (%)	33.13	23.75	13.75	13.75	***	$p = 0.139$	$p = 0.258$
	Skip after 1 or 2, %	66.04	65.79	59.09	68.18	$p = 0.793$	$p = 0.730$	$p = 0.600$
	Skip distance	2.36	1.58	1.14	1.18	***	*	$p = 0.064$
	Regression, %	70.63	60.63	48.13	50.00	***	$p = 0.272$	$p = 0.094$
	Regression to 1 or 2, %	89.38	89.69	89.61	90.00	$p = 0.929$	$p = 0.914$	$p = 0.989$
	Regression distance	2.29	1.89	1.36	1.33	***	$p = 0.075$	$p = 0.324$
Change of scan direction	ScanDown	0.79	0.66	0.59	0.65	*	$p = 0.563$	$p = 0.606$
	ScanUp	0.61	0.50	0.29	0.41	***	$p = 0.953$	$p = 0.153$
Maximum gaze position	Rank	4.28	3.59	3.28	3.06	***	**	$p = 0.195$
Trackback	Count	2.08	1.91	1.16	1.43	***	$p = 0.579$	*
Scan within link 3	In mixed strategy, %	41.38	56.70	69.05	70.53	***	$p = 0.073$	$p = 0.264$

Note. Nav denotes navigational task, Info denotes informational task, and SRP denotes search results page.
*Significant at 0.05 level. **Significant at 0.01 level. ***Significant at 0.001 level.

Search Behavior

In terms of search performance, we could find no significant difference across the screen sizes, except for the task completion duration. In this subsection, we investigate differences in users' search behavior between the large and small sizes, and discuss possible implications for their different search behaviors. Furthermore, we investigate the effect on search speed of some of the differences in search behavior. The finding suggests that subjects need to expend more effort to extract information with fewer eye movements on the small screen than on the large screen.

Fixation duration. Applying ANOVA to the total fixation duration in AOIs on search results pages (not on web documents), Table 3 showed significant differences caused by screen size ($F_{(1,605)} = 21.58$, $p < 0.001$) and task type ($F_{(1,605)} =17.81$, $p < 0.001$). This task type effect indicates that the mean fixation duration for the informational tasks was longer than for the navigational tasks on the screen sizes: 3.06 s and 2.95 s (per task) on the large screen and 4.01 s and 3.21 s (per task) on the small screen. If we connect the effect of screen size to the search speed results, we can see that users exhibited longer fixation durations per task on the small screen. To investigate this difference in detail, we calculated the mean fixation duration per link by minimal scanpath, and then examined its relationship to the search speed. The results are addressed in the subsection below on scanpath.

As for the analysis of search speed, the fixation duration data for each AOI do not meet the normality assumption. Thus, we applied the log-transformation, before using ANOVA. There is a significant effect due to screen size ($F_{(1,605)} = 23.97$, $p < 0.001$). Figure 4 describes the back-transformed fixation duration with errors on each AOI. After the first AOI, the fixation durations for each AOI decrease sharply on both screens, similar to the findings of Joachims et al. (2005) and Granka et al. (2004). However, the fixation duration on the first AOI on the small screen (1.27 s) is significantly longer than that on the large screen (0.90 s) (standard error of the difference: $SED = 0.00806$, before the back-transformation), whereas the durations on all of the other AOIs have no significant differences. This indicates that longer fixation durations on the small screen are due to AOI 1. Although the durations on other AOIs are not significantly different, Figure 4 suggests that users spent more time looking at AOIs 1–3 on the small screen than on the large screen. From AOI 4, the fixation durations seem

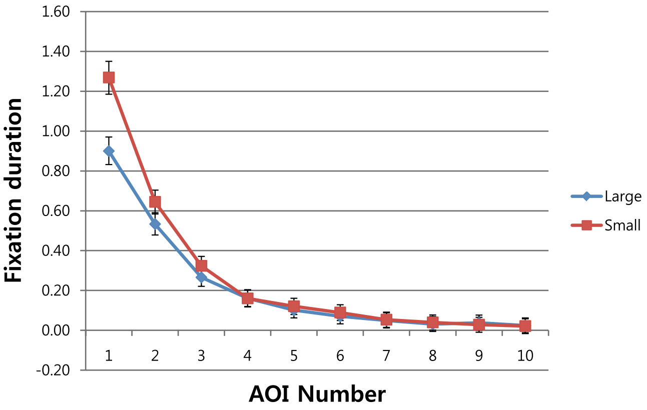


FIG. 4. Mean fixation duration on each AOI (s/user/task). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

similar for each AOI on both screens. Because users knew that the search results were from the same search engine, we cannot say that this was caused by different degrees of bias across screen sizes. Instead, this may be interpreted as an effect of the scroll function on the small screen, as suggested by Granka et al. (2004), although the decrease in fixation duration on the large screen seems to be correlated with rank order, as there is no scroll function on these screens.

Click pattern. Using a GLM with binomial distribution, a significant difference can be observed in the click pattern between task types ($\chi^2 = 14.14$, $df = 1$, $p < 0.001$), but there is no significant effect due to screen size. The click patterns in Table 3 indicate that the top three links were strongly selected on both screens and both task types (lowest: 83.1%). Although most of the relevant links were located in the top three, this is a very high proportion, even allowing for most of the fixation durations relating to these links. According to the fixation duration results, the fixation duration on the first AOI was different for each screen. However, this does not lead to more clicks on link 1 on the small screen (about 57% on both screens). This can be interpreted as meaning that users are strongly biased toward the rank orders provided by Google, as found in previous work (Guan & Cutrell, 2007; Joachims et al., 2005). The proportion of subjects choosing ranks 1–3 for navigational tasks was higher than that for informational tasks on both screen types. This is because most navigational tasks have relevant links in AOI 1, whereas the possible relevant links for informational tasks were generally located in AOIs 1–3.

Scanpath. We extracted the lengths of the minimal scanpath and compressed sequence for each task. Based on a GLM with Poisson distribution, neither screen size nor task type had a significant effect on the minimal scanpath or compressed sequence. This is consistent with findings by Lorigo et al. (2006). Table 3 shows that the minimal scanpath and compressed sequence lengths are slightly higher on the large screen and for informational tasks. However, the differences induced by the above behaviors are not statistically significant.

We next considered the difference between the compressed sequence length and the minimal scanpath length, which represents how many times users visit the same links. We found that there was no significant difference with respect to task type or screen size (GLM with Poisson distribution): The mean lengths are between 1.65 and 1.76. This indicates that users not only visited a similar number of links on both screen sizes and task types, but also revisited the same links to obtain information.

The most important finding from our study of scanning sequences is the fixation duration per link, as defined in the earlier in the subsection, Fixation duration. Based on ANOVA, there are significant effects due to screen size ($F_{(1,605)} = 64.14$, $p < 0.001$) and task type ($F_{(1,605)} = 6.04$, $p < 0.05$). When we calculated the mean fixation duration per link, we found that users needed more time to acquire information on each link when using the small screen (see Table 3). This value can be measured by dividing the mean fixation duration per task by the minimal scanpath length of the task. The fixation duration per link suggests that users had more difficulty obtaining information from the result links on the small screen (Just & Carpenter, 1976; Rayner, 1998).

To further analyze these findings, we investigated the relationship between fixation duration per link and search speed. As the time taken until the first click is not significant, but the fixation duration per link is significantly different on the different screen sizes, this analysis identifies how this difference relates to changes in search speed. Figure 5 illustrates that the relationship is clearly positive, and the slopes are similar on both screens. However, although the fixation duration per link increases with the time taken to the first click on both screens, it is significantly higher (about 0.16 s) for the small screen. This suggests that, no matter how much time subjects took to make their first click on the search results page, they normally had more difficulty extracting information on the small screen.

Scanning direction. We also examined users' scanning direction in terms of the patterns to investigate differences in how users scan links on the two screens: the complete (if the user fixated all of the links beyond a selection), linear (if the minimal path is monotonically increasing), and strictly linear (if the compressed sequence is monotonically increasing with no skips or regressions). We adopted a GLM with binomial distribution for the scanning direction data. The complete pattern exhibited a significant difference according to screen size ($\chi^2 = 17.41$, $df = 1$, $p < 0.001$). Table 3 shows that users tended to scan all links above their selection on both screens (the lowest rate is 81.8%), although there was significantly less complete scanning on the large screen. Because the complete pattern is strongly related to skipping (jumping over one link), this means small screen users either rarely skipped or more frequently jumped back to the skipped link before making a selection. The skip behavior is treated in the next subsection.

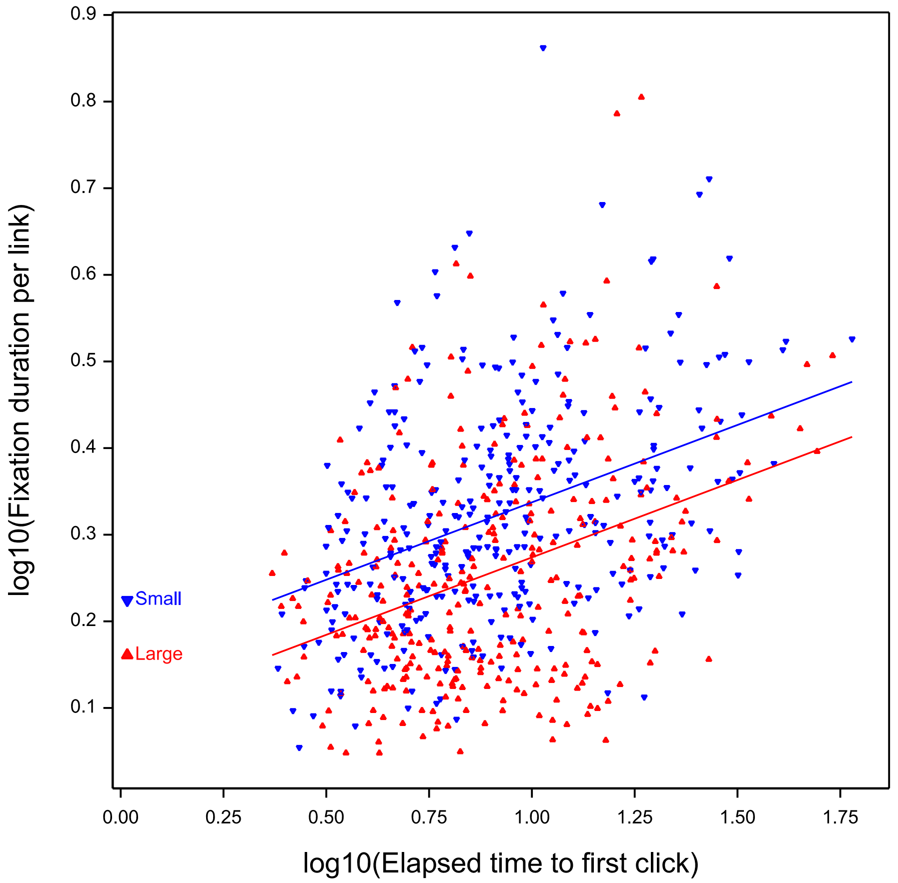


FIG. 5. Relationship between search speed (elapsed time to first click) (s) and fixation duration per link (s): The numbers on the *x*- and *y*-axes are log-transformed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

We measured two kinds of linear and strictly linear patterns. As addressed in the Measurements section, we defined linear or/ID to include the case where a user looks at and clicks only one link, because many subjects exhibited this pattern. Strictly linear or/ID was determined analogously. There were significant effects due to screen size ($\chi^2 = 30.13$, $df = 1$, $p < 0.001$), task type ($\chi^2 = 8.09$, $df = 1$, $p < 0.01$), and the interaction of these factors ($\chi^2 = 5.28$, $df = 1$, $p < 0.05$) on the linear pattern, whereas only screen size ($\chi^2 = 66.83$, $df = 1$, $p < 0.001$) had an effect on the linear or/ID pattern. This suggests that users exhibit a stronger tendency to scan links from top to bottom on the small screen and for informational tasks. The interaction effect can be explained by the relatively minor difference between task types on the large screen compared to that on the small screen. By comparing this with the results for linear or/ID, the effects of task type and interaction disappear. This is because the increase in the proportion of navigational tasks on both screens that display a linear or/ID pattern as opposed to a linear pattern is higher than that for informational tasks. This indicates that, for navigational tasks, selections on both screen types were more frequently made after looking at only one result.

Analogously, the result for strictly linear and strictly linear or/ID patterns shows the same effect for screen size (strictly linear: $\chi^2 = 12.47$, $df = 1$, $p < 0.001$ and strictly linear or/ID: $\chi^2 = 19.32$, $df = 1$, $p < 0.001$), task type (strictly linear: $\chi^2 = 7.53$, $df = 1$, $p < 0.01$), and interaction (strictly linear: $\chi^2 = 6.97$, $df = 1$, $p < 0.01$). The data for both strictly linear patterns provide evidence that the small screen leads users to scan from top to bottom without any skipping or regression. The tendency for immediate selection after looking at only one link is much stronger for navigational tasks, as for the linear or/ID pattern. This may be because it is more difficult to skip or regress on the small screen due to the need to scroll.

Skip and regression. As mentioned earlier, skipping is a significant factor in determining the proportion of complete patterns and representing how carefully users scan search results. To analyze the skip and regression data, we used a GLM with binomial distribution (except for skip and regression distances, for which we used a linear model). Table 3 shows that there were significant differences in skip rate between both screens ($\chi^2 = 21.09$, $df = 1$, $p < 0.001$). The skip rate on the small screen is about half of that on the large

DOI: 10.1002/asi

screen. Although the rate for the navigational tasks on the large screen is about 10% lower than that for informational tasks, this is not statistically significant. On both screens, users normally skipped after looking at the top ranks (SAT, about 60%) regardless of screen size or task type. In terms of skip distance, there were significant effects with respect to screen size ($\chi^2 = 15.84$, $df = 1$, $p < 0.001$) and task type ($\chi^2 = 5.92$, $df = 1$, $p < 0.05$). Although the interaction effect is not significant ($\chi^2 = 3.50$, $df = 1$, $p = 0.064$), the main difference in mean skip distance was for informational tasks on the large screen: This is more than twice the value for informational tasks on the small screen, despite there being little difference between screen types for navigational tasks. There was a significant difference in regression rate according to screen size ($\chi^2 = 19.00$, $df = 1$, $p < 0.001$). Similar to the results for skip rate, the rate of regression on the large screen was also higher than that on the small screen. The RTT proportions on both screens show that users clearly displayed a strong pattern of returning to the top ranks after their first regression. (the lowest rate is 89.38%). The regression distance was only affected by screen size ($\chi^2 = 30.09$, $df = 1$, $p < 0.001$). Users exhibited a longer regression distance on the large screen for both task types.

From our results regarding skips and regression, it is clear that subjects skipped and regressed less frequently, and a smaller distance, on the small screen. This seems to be due to the relative difficulty in skipping and getting several links back using the scroll function, or less display of items. These findings indicate that users exhibited a narrower scanning pattern over all search results on the small screen, although the minimal scanpath results indicate that the number of links examined was similar on both screens. Furthermore, compared with the results of Lorigo et al. (2006), we found a lower skip rate and a high proportion of SAT and RTT, meaning that users normally followed the rank order presented by the Google mobile search engine. This may also imply that the current search engine provides a better rank order than in the past.

Change of scan direction: ScanUp and ScanDown. As a supplement to skips and regression, ScanUp and ScanDown show how often users changed scanning direction, possibly representing hesitation during tasks. Based on a GLM with Poisson distribution, only the screen size had a significant effect on ScanUp ($\chi^2 = 23.01$, $df = 1$, $p < 0.001$) and ScanDown ($\chi^2 = 5.83$, $df = 1$, $p < 0.05$). Users showed low ScanUp and ScanDown counts on both screens: All mean values are less than 1. This implies that subjects did not generally scan over two links in the same direction, or they looked at and selected only one link. However, both the ScanUp and ScanDown behaviors were less pronounced on the small screen. In addition, although there is no significant interaction effect for either behavior, the effects of screen size seem to come from the difference between informational tasks on both screens (ScanDown: 0.79 vs. 0.59 and ScanUp: 0.61 vs. 0.29), because the difference between navigational tasks was small. Overall, this result indicates that participants changed the scan direction less in finding relevant links on the small screen, even if they looked at a similar number of links.

TABLE 4. Scroll rate, and click pattern after a scroll on the small screen.

	Info	Nav	Total
Scroll count	42	38	80
Scroll rate, %	26.3	23.8	25.0
Click links 1–3	23	24	47
Click links 1–3 rate, %	54.8	63.2	58.8

Note. Nav denotes navigational task and Info denotes informational task.

Maximum gaze position and scroll. The maximum gaze position (MGP) is useful for investigating how far from the top rank users look, and is also a prime determinant in using the scroll function on the small screen. For MGP data, we used a GLM with Poisson distribution. There are significant differences according to screen size ($\chi^2 = 26.46$, $df = 1$, $p < 0.001$) and task type ($\chi^2 = 9.26$, $df = 1$, $p < 0.01$). The MGPs in Table 3 show that users tended to look at lower ranked results on the large screen. This can be explained by the relation to the other results in this study for the scanpath and the skip and regression rates: Because users visit/revisit similar numbers of links, the great frequency and distance of skipping and regression on the large screen means that users looked farther down the results page. The MGPs for informational tasks were higher than for navigational tasks on both screens.

In addition, we can determine whether the scroll function was employed on the small screen, as the MGP will be greater than 3. The results indicate that users rarely scrolled to look down past link 3 on the small screen (MGPs of 3.28 for informational and 3.06 for navigational tasks). Table 4 describes how often users scrolled, and their click decision when using the scroll function. Only 25% of users required the scroll function, and a GLM with binomial distribution shows no significant effects by task type ($\chi^2 = 0.27$, $df = 1$, $p = 0.606$). In about 59% of the tasks, users clicked a link in the top three after scrolling to look at lower ranked results. This tendency contributes to the 88% of clicks on links 1–3 recorded over both task types. In addition, this result implies that users have a strong bias toward the rank order on a small screen, despite the need to scroll to return to the top links.

Search Strategy

We examined scanning strategies using the classification of Klöckner et al. (2004) for the initial pages of search results. The depth-first strategy (users follow a promising link immediately), the breadth-first strategy (users inspect all links exhaustively before clicking), and the mixed strategy (users read ahead, but to a smaller extent) are useful abstractions of users' decision patterns, as discussed in the Measurements section. We adopted this distinction when

TABLE 5. Choice of scanning strategy on both screen sizes and user self-assessments.

	Large			Small		
	DEP	MIX	BRD	DEP	MIX	BRD
Total	116	184	20	131	179	10
%	36	58	6	41	56	3
User self-assessments	53	47	0	69	31	0

Note. DEP denotes depth-first, MIX denotes mixed, and BRD denotes breadth-first strategy.

analyzing our data. Table 5 shows the total count and proportion of participants' scanning behaviors, differentiated by the three kinds of strategy recognized by this approach. The table shows that subjects tended to use the depth-first strategy slightly less on the large screen than on the small screen (36% vs. 41%). In contrast, on the large screen, the breadth-first strategy was used twice as often as on the small screen. However, the distribution of strategies is not significantly different across screen sizes ($\chi^2 = 4.31$, $df = 2$, $p = 0.116$). The proportion of times the mixed strategy was employed is almost the same on both screen sizes. We believe that the reason we found no significant difference between search strategies on different screen sizes may be hidden in the use of the mixed strategy. This is because the mixed strategy has too broad a definition, whereas the other two strategies are clearly defined. For example, if a user looks at links 1–3 before clicking link 1, and another scans all the way to rank 9 before clicking link 1, both are considered to be employing the mixed strategy.

To examine the detailed behavior of the mixed strategy, we adopted two metrics: trackback and how often users scanned within the top three links. First, we define "trackback" as the difference in ranks between the selected link and the farthest link observed. This allows us to scrutinize differences within the mixed strategy; the higher the trackback, the greater the extent to which links are observed. Although we investigated several search behaviors, trackback is unique in that it summarizes the amount of additional effort users make before selecting a link.

Using a GLM with Poisson distribution, we found a highly significant effect on trackback due to screen size ($\chi^2 = 48.06$, $df = 1$, $p < 0.001$) and interaction ($\chi^2 = 5.44$, $df = 1$, $p < 0.05$) across all users. Trackback values were higher across both task types on the large screen. In particular, the trackback value for informational tasks is almost double that on the small screen.

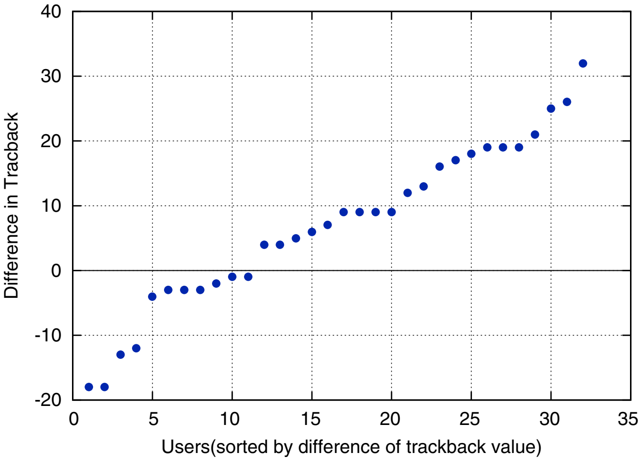


FIG. 6. Distribution of difference in trackback between both sizes of screen. Points above the *x*-axis represent higher trackback on the large screen. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

To examine the change in trackback value from large to small screens, we calculated the difference in each participant's trackback value. The difference for each user is the sum of their trackback values on the large screen minus that on the small screen. Figure 6 illustrates the difference for each participant. Points above the *x*-axis represent a higher trackback value (more looking ahead) on the large screen and points below the *x*-axis represent higher trackback on the small screen. Twenty-one users have higher trackback on the large screen whereas only 11 have higher trackback on the small screen. Clearly, the trackback value is normally higher on the large screen.

To analyze the relation between trackback and search speed, we investigated the difference in time spent examining the search results and the trackback value on both screens. Figure 7 shows that there is a positive relationship for both screens. When the trackback value is zero, there is a little difference in the time taken until the first click. However, as the trackback value increases, the difference becomes significantly larger. This result suggests that the trackback value on the small screen affects the time spent more than on the large screen. An explanation for this is that users need more time to return to the top ranks using the scroll function on the small screen, whereas they could reach the first selections using only eye movement with the large screen. In other words, this is one of the reasons why there was no significant difference in time spent selecting links, although trackback values were lower on the small screen.

Next, because users need to scroll to look below link 3 on the small screen, we analyzed how often they scan beyond the top three links when using the mixed strategy. Based on a GLM with binomial distribution, there was a significant effect due to screen size ($\chi^2 = 14.88$, $df = 1$, $p < 0.001$). Among 179 tasks in the mixed strategy was used on the small screen (see Table 5), 70% of them exhibited users' scanning only within the top three links, whereas the equivalent proportion on the large screen is some 20% lower. This means users on the small screen were more likely to exhibit a mixed scanning strategy than on the large screen when they only looked at links 1–3 without using the scroll function (e.g., users look through links 1–3 and then click link 1 or 2). In other words, users scanned fewer links on the small screen, despite the same proportion following the mixed strategy.

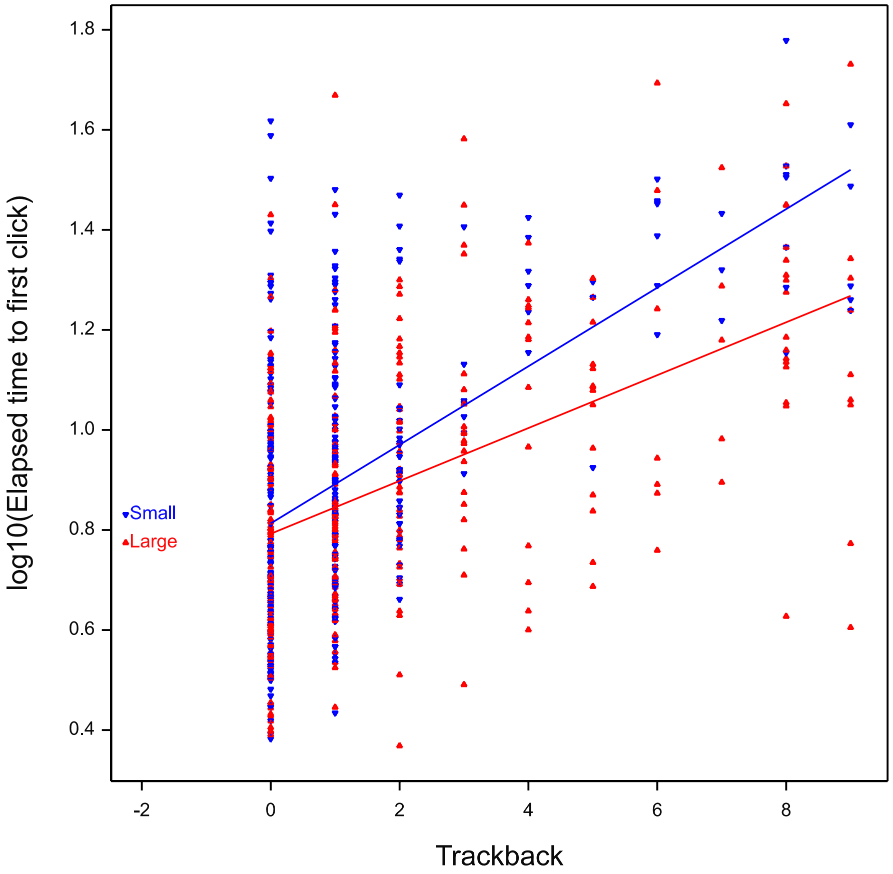


FIG. 7. Relationship between search speed (elapsed time to first click) (s) and trackback values: The numbers on the *y*-axis are log-transformed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Questionnaire

In the postexperiment questionnaire, subjects were asked several questions. All participants responded that searching on the large screen was more convenient, and about 72% of them thought the answers to each task were easy to obtain. They all considered themselves frequent and expert web searchers. In addition, their favorite search engine is Google. The main analysis of the questionnaire comes from comparing user self-assessments with the actual search strategy we observed. After the experiments, we asked subjects about their self-assessed strategy for each sized screen. The results are significantly different. None of the participants replied that they tried to scan all links of the search results as the breadth-first strategy on either screen, whereas the results show users in 6% of tasks used this strategy on the large screen and in 3% of tasks used it on the small screen (see Table 5). On the large screen, 53% of subjects reported that they had scanned the links to find only one relevant answer, and 69% reported this on the small screen. However, the actual data suggest about 17% fewer followed this strategy on the large screen and 25% fewer employed this approach on the small screen. Instead of the depth-first strategy, the mixed strategy was more often applied on both sized screens. This is because users generally intended to look for the answer quickly, but hesitated a little when they were not sure of the correct answer.

A more detailed analysis of the user self-assessments on each screen (see Appendix B1) shows that the largest proportion claimed to apply a depth-first strategy (43.75%) on both screen sizes, followed by “MIX to DEP” (25%), “MIX to MIX” (21.88%), and “DEP to MIX” (9.38%) on the large and small screens, respectively (“MIX to DEP” indicates that the user claimed a mixed strategy on the large screen and a depth-first strategy on the small screen).

We believed there would be some interesting insights when we divided the user self-assessments along task types, screen sizes, and actual strategy. However, because of small sample numbers as a result of distributing 32 users over 12 categories, the data do not seem to provide any detailed implicit meaning. We need to collect more data for the analysis.

General Discussion

We have discussed each result for search performance, behavior, and strategy, and discussed the meaning of both the explicit and implicit results. In this subsection, we summarize the discussions.

On the small screen, there were significant differences in some search behaviors, which we must consider. First, we found that it took considerably more effort for participants to gain information from each link to search results on the small screen. The relationship between the time taken to first click and the effort to extract information per link also supports this assertion. Second, from eye movement, we observed that users on the small screen scanned the search results narrowly with fewer skips and regressions as well as less frequent changes in scan direction, and exhibited a stronger tendency for reading from top to bottom. Users scanned deeper down with more skips and regressions on the large screen, despite looking at a similar number of links on both screens. In addition to the aggregate summary of search behavior, the search strategy also indicates that subjects observed the links broadly on the search results pages on the large screen. Although the difference in the proportion of search strategies on each screen was not statistically significant, subjects looked over more links before making their selection, as well as scanning below the top three links more often, on the larger screen.

However, we found several similar results in terms of search performance and behavior on both screens, as well as for both task types. First, in terms of search performance, although different amounts of time were spent viewing web documents after the first selection on the results page, no significant difference in time taken until the first click on the search results pages was observed, and users exhibited similar success rates in finding correct answers, despite the poor display for search results on the small screen. Subjects exhibited no difference in several search behaviors across both screens. The number of visited/revisited links was almost the same on both screens. Moreover, the patterns of regression to top links showed a strong bias toward the rank order provided by the search engine as it was similar to the click pattern.

Finally, small-screen users do need to concentrate more on the search results to extract information, despite displaying fewer eye movements. However, except for web documents after search results pages in informational tasks, they do not require more time to search. They also have a similar success rate of finding a correct answer. These results suggest that we need to consider these differences in search behavior and strategy for the presentation design for the search results pages, and web documents need to be adjusted to give better task completion duration, especially for informational searches, for small devices.

Limitations, Conclusions, and Future Work

In this section, we address the limitations of our experiment, and suggest some ideas for designing the presentation of search results pages. We also describe plans for future work to establish a possible presentation model for search engines for small devices.

Limitations

It is important to acknowledge that our results have several limitations. First, our results cannot cover all search performance and search behaviors, because we obtained our data from a particular user group and used a particular search engine. Another limitation is that the search environment was not perfectly natural during the experiment. Although it was designed to record eye movement in detail, users needed to put their chin on a chin rest, and they could not retype a query with different keywords. Lastly, our results will not be exactly the same as users' search behavior on actual mobile devices, because we focused only on the influence of screen size, without considering other factors such as mobility or a touch screen. We note that the results in user performance and behavior in web searches on actual handheld devices might be different if subjects used a finger for scrolling instead of a mouse or if data are recorded while the subjects are moving. We expect that the search speed will decrease and that it will be difficult to record the fixations due to the disturbance caused by the finger.

Conclusions and Future Work

In this article, we investigated the effects of screen size and task type on search performance and behavior using an eye-tracking method. In addition, we examined the relationships between search speed and some major search behaviors.

Our results suggest that although the time spent on web documents with a small versus a large screen is significantly different, users take a similar time taken to the first click on either screen. On the small screen, they have more difficulty extracting information from search results, despite their eye movement being narrower, which is a product of the scroll function on the small screen. With the limitation of less visible content on handheld devices, it does not seem that we can improve users' ability to extract information from search results pages. However, we believe that users may enjoy better search speed while maintaining their search accuracy if a presentation design could consider both the narrower eye movement and scroll function on the small screen.

As a consequence, bearing in mind the limitations of our study, we propose several ideas that possibly will be helpful for improving the search results page on small devices. First, many web providers have begun to support displaying mobile-optimized or RWD (responsive web design) web documents, and such pages are suited for the screens of small devices. Therefore, *providing a small mark indicating that an item on a search results page links to a mobile-optimized page instead of a full-size page for desktops* would contribute to reducing the user time cost in web documents, especially for informational tasks, as suggested by Jones et al. (2003).

Second, because screens in mobile devices do not present as many search results as desktop devices, and show lower search accuracy for informational tasks, we expect that *simplifying the search results with rich content* by manipulating elements such as the title or snippets may reduce the time taken to first click, as well as the task completion duration. Cutrell and Guan (2007) suggested that a longer snippet length leads to a shorter task completion time as well as better search accuracy for informational tasks, whereas the opposite result was observed for navigational tasks. In addition, our users exhibited a strong bias in relation to the rank order, and the relevant links for navigational tasks were almost located at rank 1. Therefore, one possible presentation design is to reduce the snippets to display only one row for navigational tasks, and, if necessary, allow users to extend the snippets via an extension button for informational tasks, instead of having many links on the first page of the search results.

Third, for less scrolling, as suggested by Jones et al. (1999), *embedding buttons on the search interface for Page Up and Page Down* may be useful, especially for informational tasks, to help faster and wider eye movement beyond the page break without the scroll function. Therefore, our future work will test these ideas to find the optimized presentation model for search results pages, in order to achieve better efficiency on mobile devices.

References

Aula, A., Majaranta, P., & Räihä, K. (2005). Eye-tracking reveals the personal styles for search result evaluation. Lecture Notes in Computer Science, 3585, 1058–1061.

Biedert, R., Dengel, A., Buscher, G., & Vartan, A. (2012). Reading and estimating gaze on smart phones. In Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12) (pp. 385–388). New York: ACM Press.

Broder, A. (2002). A taxonomy of web search. ACM SIGIR Forum, 36(2), 3–10.

Buscher, G., Cutrell, E., & Morris, M.R. (2009). What do you see when you're surfing?: Using eye tracking to predict salient regions of web pages. In Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09) (pp. 21–30). New York: ACM Press.

Buscher, G., Dumais, S., & Cutrell, E. (2010). The good, the bad, and the random: An eye-tracking study of ad quality in web search. In Proceedings of the 33rd international ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10) (pp. 42–49). New York: ACM Press.

Buscher, G., White, R., Dumais, S., & Huang, J. (2012). Large-scale analysis of individual and task differences in search result page examination strategies. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12) (pp. 373–382). New York: ACM Press.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. Information Processing & Management, 31(2), 191–213.

Cutrell, E., & Guan, Z. (2007). 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 (CHI '01) (pp. 407–416). New York: ACM Press.

Drewes, H., De Luca, A., & Schmidt, A. (2007). Eye-gaze interaction for mobile phones. In Proceedings of the Fourth International Conference on Mobile Technology, Applications, and Systems and the First International Symposium on Computer Human Interaction in Mobile Technology (Mobility '07) (pp. 364–371). New York: ACM Press.

Dumais, S., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for web search. In Proceedings of the Third Symposium on Information Interaction in Context (IIiX '10) (pp. 185–194). New York: ACM Press.

Findlater, L., & McGrenere, J. (2008). Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08) (pp. 1247–1256). New York: ACM Press.

Goldberg, J.H., & Kotval, X.P. (1999). Computer interface evaluation using eye movements: Methods and constructs. International Journal of Industrial Ergonomics, 24(6), 631–645.

Granka, L.A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04) (pp. 478–479). New York: ACM Press.

Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07) (pp. 417–420). New York: ACM Press.

Hotchkiss, G., Alston, S., & Edwards, G. (2005, June). Eye tracking study: An in depth look at interactions with Google using eye tracking methodology (Research white paper). Enquiro Search Solutions Inc. Retrieved from http://csi.ufs.ac.za/resres/files/hotchkiss.pdf

Jacob, R.J., & Karn, K.S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. Mind; a Quarterly Review of Psychology and Philosophy, 2(3), 4.

Jansen, B., & Spink, A. (2006). How are we searching the World Wide Web? a comparison of nine search engine transaction logs. Information Processing & Management, 42(1), 248–263.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., & Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05) (pp. 154–161). New York: ACM Press.

Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., & Buchanan, G. (1999). Improving web interaction on small screen displays. Computer Networks, 31, 1129–1137.

Jones, M., Buchanan, G., & Thimbleby, H. (2003). Improving web search on small screen devices. Interacting With Computers, 15(4), 479–495.

Just, M.A., & Carpenter, P.A. (1976). Eye fixations and cognitive processes. Cognitive Psychology, 8(4), 441–480.

Kelly, D. (2006). Measuring online information seeking context, part 2: Findings and discussion. Journal of the American Society for Information Science and Technology, 57(14), 1862–1874.

Kim, J., Thomas, P., Sankaranarayana, R., & Gedeon, T. (2012). Comparing scanning behaviour in web search on small and large screens. In Proceedings of the 17th Australasian Document Computing Symposium (ADCS '12) (pp. 25–30). New York: ACM Press.

Klöckner, K., Wirschum, N., & Jameson, A. (2004). Depth- and breadth-first processing of search result lists. In CHI'04 Extended Abstracts on Human Factors in Computing Systems (p. 1539). New York: ACM Press.

Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2006). The influence of task and gender on search and evaluation behavior using Google. Information Processing & Management, 42(4), 1123–1131.

McCullagh, P., & Nelder, J.A. (1989). Generalized linear models (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Nagamatsu, T., Yamamoto, M., & Sato, H. (2010). MobiGaze: Development of a gaze interface for handheld mobile devices. In CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10) (pp. 3349–3354). New York: ACM Press.

Nielsen, J. (2006, April 17). F-shaped pattern for reading web content. Retrieved from http://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

Palmquist, R.A., & Kim, K.-S. (2000). Cognitive style and on-line database search experience as predictors of web search performance. Journal of the American Society for Information Science, 51(6), 558–566.

Poole, A., & Ball, L.J. (2006). Eye tracking in HCI and usability research. In C. Ghaoudi (Ed.), Encyclopedia of human computer interaction (pp. 211–219). Hershey, PA/London, UK: Idea Group Reference.

Poole, A., Ball, L.J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition. In Proceedings of HCI 2004: People and Computers XVIII—Design for Life (pp. 363–378). London: Springer.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62(8), 1457–1506.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1998). Analysis of a very large Altavista query log (SRC Technical Note 1998-014). Palo Alto, CA: Systems Research Center, Compaq Computer Corporation.

Teevan, J., Alvarado, C., Ackerman, M., & Karger, D. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’04) (pp. 415–422). New York: ACM Press.

van Schaik, P., & Ling, J. (2001). The effects of frame layout and differential background contrast on visual search performance in web pages. Interacting With Computers, 13(5), 513–525.

VSN International. (2012). GenStat for Windows 15th edition. Hemel Hempstead, UK: VSN International, http://GenStat.co.uk

Appendix A

Table A1 lists the 20 tasks given to the participants. They were derived from those of Dumais et al. (2010), edited for our location.

TABLE A1. Full task descriptions and queries.

Task description	Initial task query	Task type
Find the webpage where you can apply for a personal checking account on the ANZ Bank website.	ANZ bank new account	Nav
Find the mortgage calculator on the Commonwealth Bank website where you can calculate mortgage rates for financing a new home.	mortgage rates calculator commonwealth bank	Nav
Go to the homepage of the Canberra Cavalry baseball team.	Canberra cavalry baseball	Nav
A friend of yours would like to buy some new golf clubs. Go to the official Drummond homepage.	buy clubs at drummond	Nav
Go to the official product overview page for Sony camcorders (i.e., on the Sony website).	sony camcorder	Nav
You bought a laptop from Dell and something doesn’t work as expected. Find the page for Dell technical support.	dell laptop technical support	Nav
You are interested in shoes from Nike. Go to NikeStore on the official Nike homepage.	nike shoes australia	Nav
A friend is sick and shows a couple of different symptoms. It’s nothing serious but you want to help find out what it is. Find the symptom checker webpage of WebMD.	symptom checker web md	Nav
Find the official homepage of the Canberra casino and hotel in Canberra.	Canberra Casino	Nav
Find the official Porsche website that shows Model 911 overview for Australia.	porsche australia	Nav
Find a contact number of a rental agency where you can rent a stretch limousine version of a Hummer in Australia.	rent a stretch limo hummer	Info
What is the address of the Commonwealth Bank’s headquarters (city and street)?	commonwealth bank headquarters	Info
What is the standard length of a cue used for playing billiards?	billiard cue size	Info
In what year was the Australian University established?	Australian national university history	Info
How much optical zoom does the compact digital camera Sony Cyber-Shot W530 have? None; 3x; 4x; 100x; ..?	sony cyber shot W530	Info
The new iPad 2 is out just a few months ago. In what colours can you get it (the colour of itself—not the colour of additional cases for it)?	buy ipad 2 colors	Info
How many guest rooms does the Novotel in Canberra have?	Canberra Novotel rooms	Info
The Sydney Light Rail Pass is a ticket that lets you visit many of Sydney’s sights without having to buy separate tickets each time. How much does a weekly Pass cost?	Sydney light rail fare	Info
How many spikes are in the crown of the Statue of Liberty?	statue of liberty crown spikes	Info
Find the address of an official Audi dealer near Canberra.	audi dealers Canberra	Info

Note. Nav denotes navigational task and Info denotes informational task.

Appendix B

Table B1 lists a detailed analysis of the user self-assessments and actual strategies on each screen.

TABLE B1. Detailed results of comparing user self-assessments and actual data (%).

				Large (count/user)		Small (count/user)	
Changes of strategy	Count	Rate	Strategy	Info	Nav	Info	Nav
DEP to DEP	14	43.75	DEP	31.4	37.1	47.1	34.3
			MIX	62.7	58.6	48.6	61.4
			BRD	5.7	4.3	4.3	4.3
DEP to MIX	3	9.38	DEP	53.3	53.3	46.7	46.7
			MIX	46.7	40.0	43.3	53.3
			BRD	0.0	6.7	0.0	0.0
MIX to DEP	8	25.00	DEP	35.0	25.0	42.5	37.5
			MIX	55.0	70.0	57.5	57.5
			BRD	10.0	5.0	0.0	5.0
MIX to MIX	7	21.88	DEP	48.6	31.4	40.0	40.0
			MIX	40.0	62.9	51.4	57.1
			BRD	11.4	5.7	8.6	2.9

Note. DEP denotes depth-first, MIX denotes mixed, BRD denotes breadth-first strategy, Nav denotes navigational task and Info denotes informational task.

DOI: 10.1002/asi