EFFECTS OF DEVICE SCREEN SIZE ON ONLINE INFORMATION SEARCH QUALITY & EFFICIENCY

A Thesis
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Master of Science

by
Martina Balestra

January 2015
ABSTRACT

With recent advances in mobile technologies and the growing ubiquity of wireless network accessibility, online information search tasks are now being conducted on mobile devices with a broad range of screen form factors. Screen size is particularly variable among devices, though its impact on search efficiency and quality is unknown. This study investigates the relationship between device screen size and users’ information search efficiency and quality. Thirty-six participants were tested in a variety of closed informational search tasks on three screen sizes corresponding to the Apple iPhone, iPad, and Macbook Air (13”). Although it was hypothesized that small screen size would detrimentally impact web search performance, analysis of results shows that informational search tasks were not significantly affected by screen size for measures of time on task, answer correctness, or perceived confidence or effort. The implications of these findings for mobile web searching are discussed.
BIOGRAPHICAL SKETCH

Martina Balestra received her bachelor’s degree in Mechanical Engineering from the Olin College of Engineering in Needham, Massachusetts in 2010. In the last year of college she discovered the field of Human Factors engineering, and pursued her interest as a Human Factors specialist at the MITRE Corporation upon graduating. After two years, she realized that she was interested in conducting more robust research in and found herself in the Design and Environmental Department at Cornell University pursuing an M.S. Upon completing her degree at Cornell, Martina hopes to contribute to the field as a PhD student in Human Computer Interaction at New York University.
ACKNOWLEDGEMENTS

I would first like to acknowledge my advisor, Alan Hedge, for leading me through this experience with a sense of humor. Our many thought-provoking conversations and shared curiosity have inspired and motivated me to pursue these topics in academia and beyond. I would also like to thank my minor advisor, Tarleton Gillespie, and various other teachers and mentors far and wide for their patience and honesty. Finally, I will be eternally grateful to the friends, family, and loved ones who have inspired and cared for me throughout this experience and countless others.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>III</td>
</tr>
<tr>
<td>BIOGRAPHICAL SKETCH</td>
<td>IV</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>V</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>VII</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>VIII</td>
</tr>
<tr>
<td>CHAPTER 1 - INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 - LITERATURE REVIEW</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER 3 - RESEARCH OBJECTIVES</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER 4 - METHODOLOGY</td>
<td>33</td>
</tr>
<tr>
<td>CHAPTER 5 - RESULTS</td>
<td>41</td>
</tr>
<tr>
<td>CHAPTER 6 - DISCUSSION</td>
<td>57</td>
</tr>
<tr>
<td>CHAPTER 7 - CONCLUSION</td>
<td>65</td>
</tr>
<tr>
<td>CHAPTER 8 - LIMITATIONS</td>
<td>66</td>
</tr>
<tr>
<td>APPENDIX A - QUESTIONS &amp; ANSWERS</td>
<td>70</td>
</tr>
<tr>
<td>APPENDIX B - SURVEYS</td>
<td>71</td>
</tr>
<tr>
<td>APPENDIX C - STUDY RECRUITMENT MATERIALS</td>
<td>78</td>
</tr>
<tr>
<td>APPENDIX D - INFORMATION SEARCH MODELS</td>
<td>79</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>82</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1 Sample search result components ............................................................. 2
Figure 2 Google click share by device ....................................................................... 12
Figure 3 Comparison of SERPs on mobile and traditional devices ......................... 14
Figure 4 The traditional computer-based information retrieval model ..................... 17
Figure 5 Information retrieval model augmented for the web ..................................... 10
Figure 6 Factors influencing needs and information seeking behaviors .................... 15
Figure 7 Johnson & Meischke’s Information Seeking Model .................................... 17
Figure 8 Apple device screen dimensions ................................................................ 37
Figure 9 Device ownership breakdown. ..................................................................... 34
Figure 10 Phone ownership breakdown. .................................................................... 34
Figure 11 Mean Time on task by device ................................................................ 41
Figure 12 Correct answers by device ........................................................................ 42
Figure 13 Self reported confidence score by device .................................................. 43
Figure 14 Self reported effort score by device. ........................................................... 44
Figure 15 Mean time on Google result pages in seconds .......................................... 54
Figure 16 Mean time on content pages in seconds .................................................... 54
Figure 17 Mean times spent on pages ...................................................................... 55
Figure 18 Mean time scrolling and stationary on search pages ................................. 56
Figure 19 Mean times scrolling and stationary on content pages ............................... 57
Figure 20 Number of search results viewed .............................................................. 58
Figure 21 Number of results pages viewed ............................................................... 59
Figure 22 Cumulative search query length ................................................................ 60
Figure 23 Mean frequency with which participants navigated to a branch step .......... 61
Figure 24 Mean frequency with which participants navigated to a branch step .......... 53
Figure 25 Mean frequency with which links were clicked. ........................................ 55
LIST OF TABLES

Table 1 Search engine market shares ........................................... 2
Table 2 Title, url, abstract size, and number of chars. across devices ........ 15
Table 3 Hierarchical classification and definitions of user intent as expressed by Web queries ........................................... 22
Table 4 Comparison of information search behavior frameworks .............. 19
Table 5 Summary table of major studies on informational search behaviors ... 20
Table 6 Percentage of interactions classified as searching or browsing ......... 25
Table 7 Descriptive statistics on search trail features ................................ 25
Table 8 U.S. consumer smartphone market share .................................. 36
Table 9 U.S. consumer tablet market share ........................................ 36
Table 10 Demographic summary ..................................................... 33
Table 11 Representation of experiment design ...................................... 35
Table 12 Closed, informational questions presented to participants .......... 47
Table 13 Probability of a participant navigating to a website link level ........ 54
Table 14 Probability of clicking a link on the results page ......................... 64
CHAPTER 1
INTRODUCTION

Advances in mobile computing technologies and the growing ubiquity of network accessibility over the last decade have substantially expanded the context of computer use, the availability, and value of information. As asserted by Brown and Duguid (2002), “Information is seen as not only a necessary, but also a sufficient condition for learning, work, and life itself” (p. 429). Indeed, the digital revolution has shifted value from human capital in the industrial age to information capital in the digital age (Rheingold, 2003), which has led to the decentralization of the production, ownership, dissemination, and consumption of knowledge and culture (Benkler, 2006; Glaeser, 2011; Rheingold, 2003) and become a primary currency in innovation (for example, see Brown and Duguid (2002) for a description of the role of information in the foundation of Silicon Valley as a site for innovation). Over the last two decades these changes have led to a fundamental and widespread shift in the way we interact with information, and information search itself has become an intrinsic part of the way we now conceptualize consumer empowerment (Peterson & Merino, 2003) and self-efficacy (McLeod, Scheufele, & Moy, 1999).

Anytime, Anywhere Search

Information search on computers (versus other forms of conventional media) was initially limited to wired desktops in specially designated locations and contexts. By comparison, contemporary mobile device and network technologies permit “anytime, anywhere” search whereby users are able to search for information relevant to their
immediate needs and physical context using Internet enabled search engines. A number of these search engines exist, allowing users to look for information in existing web pages, images, databases, and directories. Search can be conducted through a variety of means including peer-to-peer networks (e.g. FAROO), metasearch (e.g. Blingo), based on geography (e.g. ZipLocal), or subject matter specific resources (e.g. Monster.com provides a search engine exclusively for job related information.) The most widely used search engine, Google, relies on natural language processing of queries to retrieve relevant information from publicly available resources. Google’s search engine results pages (SERPs) present users with the title of the results, a brief description or abstract, and a reference to the source (i.e. url) (figure 1).

![Sample search result components](image)

**Figure 1** Sample search result components

Table 1 demonstrates Google’s market share compared with other common search engines:

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>Market Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google search engine</td>
<td>67.5%</td>
</tr>
<tr>
<td>Microsoft sites</td>
<td>18.4%</td>
</tr>
<tr>
<td>Yahoo sites</td>
<td>10.3%</td>
</tr>
<tr>
<td>Ask network sites</td>
<td>2.4%</td>
</tr>
<tr>
<td>AOL sites</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

According to the Pew Internet & American Life Project (Duggan & Smith, 2013), 57% of all Americans (that is, 63% of all cell phone owners) use their Internet enabled cell phones (called smartphones) to go online. 11% of smartphone owners use their
computers and phones equally, and 34% *predominantly* use their smartphones to access the Internet. While the proliferation rate of tablet computers is somewhat lower than that of smartphones, one report published by the Consumer Electronics Association (2014) found that 44% of online U.S. consumers owned and used tablets by 2014. “Tablet” computers describe a variety of mobile device in which all components of the device are located in a single, hand-held unit. As is the case with smartphones, the keyboard and mouse are typically replaced with on-screen gestural manipulation. Tablets tend to differ from smartphones in size and their ability to place phone calls, however tablet users can access the Internet using the same browser applications as smartphone users. A study conducted by the digital marketing agency, The Search Agency (2014), found that mobile tablet and smartphone search is commanding an increasing portion of all Google searches (figure 2).

![Figure 2 Google click share by device (The Search Agency, 2014)](image_url)
**Alternative Search Environments**

Computing devices that are specially designed to be mobile can differ substantially from traditional desktops and laptops in both form factor and interaction paradigm. Mobile devices are usually much smaller than their traditional counterparts, and they frequently employ direct manipulation interaction paradigms (that is, users can reach out and interact directly with objects on the screen). Most standard tablets and smartphones allow users to navigate and type using their fingertips to type on on-screen keyboards that only appear when they are needed. Desktop and laptop computers generally have physical and permanent mice and keyboards. While these on-screen keyboards usually have the same layout as physical keyboards, they are generally much smaller and thus have limited functionality in the interest of space.

Figure 3 provides a visual comparison of the screen dimensions of the Apple mobile product line (particularly the iPad and iPhone) with a standard 13” Mac laptop, demonstrating how differences in screen size can impact the display of search engine results.
Figure 3 Comparison of SERPs on mobile and traditional devices, including screen dimensions for each device.

The screen size variability among devices impacts the amount and layout of information displayed in the three primary components of results on the SERP (table 2).
Table 2 Title, url, abstract size, and number of chars. across devices

<table>
<thead>
<tr>
<th></th>
<th>Title - terms (%) relative to laptop</th>
<th>url - characters (% relative to laptop)</th>
<th>Abstract - terms (% relative to laptop)</th>
<th>Google text input bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPhone</td>
<td>7.1 (97)</td>
<td>42.5 (73)</td>
<td>18.1 (71)</td>
<td>15 chars.</td>
</tr>
<tr>
<td>iPad</td>
<td>7.1 (97)</td>
<td>58.1 (100.1)</td>
<td>25.4 (99)</td>
<td>22 chars.</td>
</tr>
<tr>
<td>Mac Book Air</td>
<td>7.3 (100)</td>
<td>58.0 (100)</td>
<td>25.6 (100)</td>
<td>42 chars.</td>
</tr>
</tbody>
</table>

Though reduced screen size also implicates challenges for users’ manual dexterity and the physical manipulation of information (small icons, small keyboards, etc.), the present study aims to look exclusively at the role of screen size on differences in search performance as a function of the visual presentation of information. As efficient and high quality information search has become an important part of contemporary life, this study hopes to contribute to an understanding of the ways in which our perception and interaction with information is augmented by the devices we choose to use, and in turn pave the way for a more thorough approach to testing the “usability” of mobile devices.
CHAPTER 2
LITERATURE REVIEW

The increasing accessibility of digital information has resulted in many studies on information search since the 1980s. The most mature research tends to come from a computer science perspective. In particular, the fields of information retrieval (IR), natural language processing (NLP), and machine learning have provided a substantial amount of literature on the underlying Boolean (Lancaster & Fayen, 1973), probabilistic (Fuhr, 1992), and vector space (Raghavan & Wong, 1986) algorithms used to aggregate and rank search results. Search performance and efficiency with respect to alternative IR algorithms (Zorn, Emanoil, & Marshall, 1996), and especially the design and refinement of natural language information retrieval algorithms (Lewis and Sparck-Jones’ (Lewis & Spark Jones, 1996; Strzalkowski, 1995), have also been studied in detail. The development of natural language IR, which allows users to formulate requests using intuitive expressions from natural speech rather than a formal query language, underpin contemporary information services such as the Google search engine.

Despite an algorithmic focus in the existing literature, the need for behavioral research focusing on how humans perceive and consume information search results does not go unrecognized. Indeed, Dervin and Nilan (1986) suggested “changing the procedures by which user needs are assessed in practice, from keyword, symbol-matching, and subject-orientations to user-problematic situations” (p. 8) in the development of such studies. Tsai et al. (2010) echo this advice with more contemporaneously relevant recommendations to expand research in multimodal user interfaces and mobile media retrieval. Research on these topics, however, appears to remain sparse. Studies tend to fall into one of two categories: adoption (for example, the
Pew Internet and American Life Project which attempts to measure the evolution of the internet and its usage (Duggan & Smith, 2013), or (less frequently) empirical behavioral studies like the present one.

This literature review will first discuss the information search models that underpin any and all search processes, followed by a review of research detailing the factors that impact information search behaviors (ISB) (focusing on search in the mobile environment in particular.) The final section of this chapter will review prominent empirical studies attempting to quantify individual search behaviors using online search engines in mobile contexts where available.

*Models of Information Search*

The classic model of media-agnostic information retrieval is shown in Figure 4 and is described in van Rijsbergen’s (1979) seminal text *Information Retrieval*: “Suppose there is a store of documents and a person (user of the store) formulates a question (request or query) to which the answer is a set of documents satisfying the information need expressed by his question.” (p. 3)

![Figure 4](image.png)

*Figure 4* The traditional computer-based information retrieval model (van Rijsbergen, 1979, p. 4)
Van Rijsbergen (1979) IR model describes the process of informing a system’s user on the existence of information related to the query they submitted to the processor. The processor itself compares the query with all available files (labeled “documents” in the model) and provides some representation (the output) of potentially relevant information. The user evaluates the results in the output and either reformulates the query to elicit different results, or uses the information presented. Implicit in the model are the individual actions users must take to examine the output, and formulate feedback if necessary.

Broder (2002) extended the traditional IR model to provide one of the first (and widely adopted (as by Lorigo et al. (2006), Kang & Kim (2004), and Evans & Chi (2008)) taxonomies of web-based information search (figure 5). Broder (2002) expanded on the original conception of input to specifically call out the elements of the user’s context – task, information need, and verbal form – that inform the construction of a query.
This web-centric IR model also provides a basis for search in mobile environments because it implies a legitimate role for the search context. Both the physical environment of the user and the constraints of their digital environment impact the factors Broder (2002) has outlined as impacting query input. For example, Kamvar et al. (2009) suggest that users exhibit different information needs on different devices (“e.g. users may be more likely to search for quick factual information on mobile phones.” (p.809)) that may be influenced by the ambient context of the user.

Studies specifically looking at search in the mobile context are predominantly concerned with the designs, opportunities, and challenges of modes of input. Inspired by the expanded sensing capabilities of mobile devices, they present solutions for small screen dimensions by eliminating the need for direct manipulation of search results all
together. For example, Chang et al. (2002) presents an IR system for spoken natural language information search in response to devices’ high-quality microphones and supposedly inefficient keypad typing. Xie et al. (2008) discuss using image files, taken with the smartphone’s camera, as a search query for finding similar images, or to search for the same concept as contained in the image.

A different line of research makes the fundamental assumption that mobile information searches are predominantly interested in search results relevant to their immediate geographic and social context. For example, Tsai et al. (2010) discuss the implications of context aware information search that incorporates location-based search, mobile peer tagging, and mobile learning among other features in the presentation of results the system interprets to be most salient at the top of the list.

**Factors Effecting Search Behavior**

*Information Need.* As suggested by Tsai et al. (2010), an essential component of understanding how to adapt information to the mobile environment lies in how human behavior manifests in the mobile environment, and recalling Broder’s (2002) user centric view of information search. Broder (2002) extended traditional conceptions (see (van Rijsbergen, 1979)) of information search using Schneiderman et al.’s (1997) concept of information need: “the perceived need for information that leads to someone using an information retrieval system” (p. 12), though the concept of “information need” was coined by Robert Taylor in *The Process of Asking Questions* (1962). According to Taylor (1962), there exist four levels of information need: “(1) the actual, but unexpressed, need for information (the *visceral* need) (2) the conscious within-brain
The motivation for constructing the query can provide compelling explanations of information use, aid in predicting information usage, and improve information manipulation (Dervin & Nilan, 1986) by informing the design of more salient search engines that anticipate the expectations and limitations of the user.

**Type of Information Search.** Broder (2002) identifies three types of information searches, dependent on information need from device-agnostic transaction logs: “1. **Navigational.** The immediate intent is to reach a particular website. 2. **Informational.** The intent is to acquire some information assumed to be present on one or more web pages. 3. **Transactional.** The intent is to perform some web-mediated activity.” (p. 5, emphasis original). In a device agnostic query log analysis of 1000 randomly selected queries, Broder (2002) found that 20% of web search queries were navigational, 48% of queries were informational, and 30% were transactional. In a more recent study, Rose and Levinson (2004) found slightly different proportions - approximately 13.5% of queries were navigational, 62% were informational, and 24.5% were transactional.

Furthermore, a study comparing mobile transaction log analyses between the years 2005 and 2007 concluded that users in 2007 tended to engage in unstructured exploration of information more frequently and use less homogenous queries than the users in the 2005 data. While this study did not distinguish between the types of information search, it did provide one of the first examples of tying specific behaviors (i.e. words/characters per query, percent of queries that had at least one click,
percent of queries that had at least one “more search results” request, time to enter a query, time between receiving results and clicking on a spelling correction for a query, time between receiving results and clicking on a search result) to search objectives.

Jansen et al., (2008) used combinations of “(1) empirical studies and surveys of search engine use; (2) manual analysis of search engine transaction logs; (3) automatic classification of Web searches” (p. 1252) to develop a more formal taxonomy of intent in web search, identify characteristics of each query type, operationalize the classification to support real-world classifications, and test the effectiveness of the taxonomy against real, large transaction logs. The resulting taxonomy is shown in table 3:

<table>
<thead>
<tr>
<th>Level</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>User intent classification</td>
<td>Informational (queries meant to obtain data or information in order to address an information need, desire, or curiosity)</td>
<td>Directed (specific question)</td>
<td>Closed (deals with one topic; question with one, unambiguous answer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Undirected (specific question)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Find (locate where some real world service or product can be obtained)</td>
<td>Open (deals with two or more topics)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>List (list of candidates)</td>
<td>Advice (advice, ideas, suggestions, instructions)</td>
<td></td>
</tr>
<tr>
<td>Navigational (queries looking for a specific URL)</td>
<td>Navigation to transactional (the URL the user wants is a transactional site)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Navigation to informational Online (the URL the user wants is an informational site)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transactional (queries looking for resources that require another step to be useful)</td>
<td>Obtain (obtain a specific resource or object)</td>
<td>Off-line (the resource will be obtained off-line and may require additional actions by the user)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Download (find a file to download)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Information Environment. Knight and Spink (2008) assert that information search behaviors (ISB) arise from the characteristics and behaviors associated with a specific information environment, in addition to the information needs and individual qualities of the information seeker. It is important to note that information environment refers not only to the larger physical context of the user, but the physical characteristics of the information environment itself. Of the various factors affecting mobile information search, the impact of its environment is the least represented in the literature. This may be a symptom of the relative infancy of mobile search in general: while mobile phones with Internet accessibility have been available since 1993, mass adoption was not seen in the U.S. until the mid-2000s with the Blackberry, and even more so in 2007 with the introduction of the iPhone. Indeed, one of the most widely cited articles in the space (Jansen, Spink, & Saracevic, 1998), in which the authors describe a large scale study focusing on queries, sessions, and terms used in desktop, information search, was published in 1998, long before mobile information search became prevalent.
Figure 6 demonstrates how Wilson (1981) incorporates the concept of information environment into his model of information behavior:

Wilson (1981) asserts that “…the physical environment will have a clear effect upon the nature of some categories of tasks and upon the consequent cognitive need” though in 1981 it may have been difficult for Wilson to imagine the variety of information environments people would be using by the year 2014.

Cognitive Load Theory (CLT) (Moray, 1979) highlights the relevance of device screen size, as an element of information environment that impacts the physiological,
affective, and cognitive needs of the user, implied by Wilson (1981). CLT refers to mental workload as the interaction between the requirement of mental resources imposed by a task or system, and an individual’s ability to accommodate the required resources (Moray, 1979). Of particular relevance to online information search is the finite capacity of the visuo-spatial sketchpad, the cognitive system that manages visual information processing: “when demands of one task are high, the resources committed to that task become unavailable to a second task if it requires the same type of mental resources, …, and at the same stage of processing (i.e. cognitive versus response related)” (Gwizdka, 2010). In the context of information search, Gwizdka (2010) asserts that Cognitive Load Theory describes why information displays with more complex visual features have higher cognitive processing demands, a positive correlation between navigation path length and a subjective assessment of task difficulty, and that mental workload is lower when on-screen text was fitted to the screen (instead of requiring scrolling.)

Information Search Strategies. Understanding information search strategies and information foraging theory in particular has practical importance because, as Lorigo et al. (2006) suggest, it seeks to compare the time, opportunity, and resource “cost[s], or perceived cost[s], of reformulating a query, the cost of clicking on a suggested document link, and the cost of gazing, or reading the abstract, overtime” (p.1130). Johnson and Meischke’s (1993) suggest that information-seeking decisions are based heavily in the context of the information seeker, as seen in figure 7:
In this model Johnson and Meischke interpret utility as the quality of the information as it relates to the need of the individual. Characteristics are the qualities related to message content attributes and communication potential, “…an individual’s perception of the manner in which information is presented. This dimension relates to issues of style and comprehension” (Johnson & Meischke, 1993, p. 349). Information seeking actions include method (relates to the media channel selected for search, i.e. search engine), scope (the number of different elements in the medium that are perceived), and depth (the extent of readership of particular elements of the medium) (Johnson & Meischke, 1993, pp. 350-351).
While Johnson and Meischke’s (1993) model provide a good summary of factors that may impact search strategy decisions, it may be more useful to consider the widely cited (albeit more generalized) search strategies put forth by Pirolli and Card (1999) in attempting to understand specific behaviors.

Pirolli and Card’s (1999) theory, known as information foraging, is built on the assumption that information seekers will modify their search strategy or search environment in order to maximize their rate of gaining information. The theory is at its core a cost/benefit question in which the interface between the information seeker and the information determines the time, resource, and opportunity costs of a given information searching strategy comprised of individual actions and behaviors (Pirolli & Card, 1999).

The authors assert that an optimization model of information search strategy in a given search environment would be comprised of three components:

1. Decision assumptions, which specify the decision problem to be analyzed. Examples of such information foraging decisions include how much time to spend processing a collection of information or whether or not to pursue a particular type of information content.

2. Currency assumptions, which identify how choices are to be evaluated. Information foraging theory will assume information value as currency. Choice principles include maximization, minimization, and stability of that currency.

3. Constraint assumptions, which limit and define the relationships among decision and currency variables. These will include constraints that arise out of the task structure, interface technology, and the abilities and knowledge of a user population. Examples of constraints include the rate at which a person can navigate through” (Pirolli & Card, 1999, p. 644)

The authors conclude, as Johnson and Meischke (1993) did, that the weights applied to the various costs and assumptions are dependent on the
disposition of the seeker, and that the relationships between seeker, information, and environment are highly dynamic, constantly acting on each other.

**Search Behaviors**

Broad categorizations and characterizations of search behaviors have been found in a number of studies, most of which predate the mass adoption of smartphones. Table 4 summarizes several of these models the slightly different levels of granularity.

<table>
<thead>
<tr>
<th>Ellis (1989a; Ellis, Cox, &amp; Hall, 1993) Appendix D, Table 1</th>
<th>Shneiderman, Boyd, &amp; Croft (1997) Appendix D, Table 2</th>
<th>Kuhlthau (1991) Appendix D, Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting</td>
<td>Formulation</td>
<td>Initiation</td>
</tr>
<tr>
<td>Chaining</td>
<td>Action</td>
<td>Selection</td>
</tr>
<tr>
<td>Browsing</td>
<td>Review of results</td>
<td>Exploration</td>
</tr>
<tr>
<td>Differentiating</td>
<td></td>
<td>Formulation</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Extracting</td>
<td></td>
<td>Collection</td>
</tr>
<tr>
<td>Verifying</td>
<td></td>
<td>Presentation</td>
</tr>
<tr>
<td>Ending</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following sections will address research attempting to quantify specific informational search behaviors. Table 5 summarizes key findings from relevant studies.
Table 5 Summary tables of major studies on informational search behaviors

<table>
<thead>
<tr>
<th>Title</th>
<th>Author (s)</th>
<th>Year</th>
<th>Participants</th>
<th>Study design &amp; Methods</th>
<th>Main Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>The influence of task and gender on search and evaluation behavior using Google</td>
<td>Lorigo, Pan, Hembrooke, Joachims, Granka, &amp; Gay</td>
<td>2005</td>
<td>N=23; 18-23 y/o; 14 males, 9 females</td>
<td>Transaction log and eye-tracking analysis of 5 informational and 5 navigational search tasks on laptop using Google. ANOVA test of continuous vars., Chi square test for categorical vars.</td>
<td>60% of questions answered correctly, avg. 1.6 queries per task, avg. time on task = 46s, % time on Google result pg. per task = 40.2%</td>
</tr>
<tr>
<td>How does search behavior change as search becomes more difficult?</td>
<td>Aula, Khan, &amp; Guan</td>
<td>2010</td>
<td>N=179; 18-54 y/o;</td>
<td>Avg. 22.3 directed closed informational tasks of varying difficulty on laptop using online search engine.</td>
<td># of queries decreased as task success rate increased, harder tasks resulted in more and longer queries, spent more time on results pg.</td>
</tr>
<tr>
<td>A Large Scale Study of Wireless Search Behavior: Google Mobile Search</td>
<td>Kamvar &amp; Baluja</td>
<td>2006</td>
<td>&gt;1,000,000 mobile Google transaction logs</td>
<td>Massive transaction log analysis comparing 12-key keypad cell-phones, phones w/ QWERTY keyboards (i.e. PDAs), and conventional computers.</td>
<td>Avg. length per query = 15.5 chars., queries, # of terms, and length were categorized by topic; 1.7 clicks per query.</td>
</tr>
<tr>
<td>Computers and iPhones and mobile phones, oh my!</td>
<td>Kamvar, M.; Kellar, M.; Patel, R.; Xu, Y.</td>
<td>2009</td>
<td>&gt;100,000 queries issued over 10,000 users during a 35 day period</td>
<td>Massive transaction log analysis.</td>
<td>Query length (characters): laptop = 18.72, iPhone=18.25, mobile=15.89. Avg. # of queries / session: computer=1.94, iPhone=1.82, mobile=1.70</td>
</tr>
</tbody>
</table>

(1) Efficiency & Success – In a study conducted on conventional computers comparing search task type (i.e. informational, navigational, or transactional) using Google, and gender, Lorigo et al. (2006) found that search success rates did not differ significantly between either search type or gender, though the authors did find that informational tasks required longer to complete than navigational tasks (with significance at the 0.05 level.)
In a large study of closed informational search tasks using Google, Aula, Khan, and Guan (2010) also found that successful tasks tended to require less time (176.2s) compared with unsuccessful tasks (384.6s) (the mean time to completion for all tasks was 223.9s). The authors suggest this may be a result of users attempting alternative or less intuitive search strategies as the questions became more challenging.

(2) Browsing – In a large online study (N=179), Aula, Khan and Guan (2010) found that participants tended to spend more time on the Google SERP as the task difficulty increased (approximately 10 seconds in successful tasks and 13 seconds in unsuccessful tasks, p<0.0001). When considered as a proportion of the total time on task, participants still spent more time on the search results page as difficulty increased (approaching 50% of the time).

In a transaction log analysis of 1,000,000 search sessions, Silverstein et al. (1999) observed that only the first page of search results was viewed in 85% of queries, though an average of 1.3 screens were viewed per query overall. In a study of 437 queries, Lorigo et al. (2006) found that subjects spent 40.2% of time on the Google SERP, and a mean 26.1 seconds viewing web documents per search session (average 12.1 seconds on a single web document compared with all documents selected in a session.) Further, Kamvar and Baluja’s (2006) study comparing search on conventional computers, personal digital assistants (PDAs), and conventional cell-phones found that users spent between 29 and 36 seconds on the SERP before clicking a link. This study also used the Google search engine as the information environment, however the authors note that the mobile browser was not as advanced at that point as the desktop browser: PDAs were
typically only capable of displaying HTML (but not JavaScript), and other cell phones only displayed limited XHTML.

A number of studies using eye-tracking methodologies to understand website usage on conventional computers (Nielsen & Pernice, 2010; Lorigo, et al., 2008) shows that people tend to browse pages using an “F-pattern” in which users predominantly read the top two horizontal bars of text then vertically scan the remaining text. Indeed, Lorigo et al. (2008) found that links in the first two ranks received the majority of attention in Google search, at 76.7% and 56.6% for the first and second respectively. This literature review found no comparable studies examining eye scan path on mobile devices.

(3) Query – Jansen et al.’s (2008) study classifying search behaviors based on information search type identified the following characteristics of queries in informational search tasks using the search engine “Dogpile”:

- “Uses question words ((i.e., ‘ways to’, ‘how to’, ‘what is’, etc.);
- queries with natural language terms; queries containing informational terms (e.g. list, playlist, etc.);
- queries that were beyond the first query submitted;
- queries where the searcher viewed multiple results pages; queries length (i.e., number of terms in a query) greater than 2; and
- queries that do not meet criteria for navigational or transactional.” (p. 1261)

Though the article does not provide quantitative support for these characteristics, the authors also assert that informational search tasks tend to result in longer queries than with other types of tasks, with more refinements submitted to the database. Other device agnostic transaction log analyses tended to agree that most queries contain on average between 2.35 and 2.6 terms (Jansen, Spink, & Saracevic, 1998; Silverstein, Henzinger, Marais, & Moricz, 1999)
A Google transaction log analysis conducted by Kamvar et al. (2009) deliberately comparing query processes on conventional computers and iPhones also observed that query length was comparable on both devices (though tests for statistical significance were not discussed.) Aula et al. (2010) also found, however, that more difficult tasks required longer queries that tended to become more specific and complex over the search session.

Aula et al. (2010) found participants using an average of 4.77 terms per query, however the objective of the study was to test how search behavior changes as tasks become more difficult, and thus the increased number of queries may reflect the deliberately difficult questions used in the study. This may also account for higher query submission rates than found in other studies. Aula et al. (2010) observed higher query submission rates in successful tasks (average 4.98 queries were submitted) versus 12.41 queries in unsuccessful tasks (both relationships were significant with p-values <0.0001).

In a study comparing search patterns across computers, iPhones, and PDAs, Kamvar et al. (2009) noted that more queries tended to be submitted on computers than on iPhones per search session (1.94 and 1.84 on average, respectively). The study was conducted on a general transaction log in which task type was not controlled for, leading the authors to hypothesize that this behavior may be a result of different information needs associated with search on computers versus mobile devices. For example, they suggest that users are more likely to search “quick, factual information” (Kamvar, Kellar, Patel, & Xu, 2009, p. 809) search on mobile devices. The authors also suggest that users may be discouraged by more difficult text typing, and that “users on mobile phones may
be more likely to browse multiple results in place of issuing query refinements” (Kamvar, Kellar, Patel, & Xu, 2009, p. 809).

A transaction log analysis conducted by Church et al. (2007) observed that a large portion of query refinements did not result in a change in query length, indicating that many refinements consisted of systematically swapping out terms to refine the search results.

(4) Search Navigation – In their widely cited article *A large scale study of wireless search behavior: Google mobile search*, Kamvar and Baluja (2006) reported relatively low click rates during Google search on PDAs (1.7 clicks per query) citing that users most likely relied on information contained in the abstract of the SERP. The authors also use this finding to support their hypothesis that there is relatively little exploration in wireless search, however the study employed a data set of over 1 million search sessions that were not controlled for search type or objective.

White and Drucker’s (2007) search engine agnostic log analysis of 3,291 individuals search behavior revealed interesting interactions between searching (defined here as “a directed search involving a recognized commercial search engine” (p. 24)) and browsing behaviors (defined as “a view of a page that lies somewhere on the click path flowing away from a search results page” (White & Drucker, 2007, p. 24). White and Drucker classify actions “forward” (“where a user clicks a hyperlink to visit a page not previously visited on the search trail” (p. 24)) and “backward” (“where a user revisits a page on the trail” (p. 24)) to measure their general frequencies in online search, table 6.
Table 6 Percentage of interactions classified as searching or browsing (White & Drucker, 2007, p. 24)

<table>
<thead>
<tr>
<th>% Total Interactions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Backward-to-browse</td>
<td>21</td>
</tr>
<tr>
<td>Forward-to-search</td>
<td>21</td>
</tr>
<tr>
<td>Backward-to-search</td>
<td>8</td>
</tr>
<tr>
<td>Forward-to-browse</td>
<td>50</td>
</tr>
</tbody>
</table>

White and Drucker (2007) also provide what appears to be one of the few comprehensive accounts of browsing behavior, albeit solely conducted on the conventional computer. The following table provides descriptive statistics of search trail features presented in their paper:

Table 7 Descriptive statistics on search trail features (White & Drucker, 2007, p. 27)

<table>
<thead>
<tr>
<th>Search Trail Feature</th>
<th>Per Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>435.4 s</td>
</tr>
<tr>
<td>Number of queries</td>
<td>4.0</td>
</tr>
<tr>
<td>Number of steps</td>
<td>16.5</td>
</tr>
<tr>
<td>Number of revisits</td>
<td>4.3</td>
</tr>
<tr>
<td>Number of branches</td>
<td>3.6</td>
</tr>
<tr>
<td>Average branch length</td>
<td>4.5</td>
</tr>
</tbody>
</table>

- **Time** Amount of time spent (in seconds) on a trail.
- **Number of queries** The number of queries that were submitted during a trail.
- **Number of steps** The number of pages viewed in a trail, including all searches and revisits.
- **Number of revisits** The number of revisits to a page viewed earlier in the trail. Revisits to pages viewed previously in other trails are disregarded.
- **Number of branches** The number of times a subject revisited a previous page on the trail and then proceeds with forward motion to view another page.
- **Average branch length** The average number of steps in each branch in the trail.
The authors use this information to test hypotheses of user variability in search processes, however it also sets a precedent for additional search behavior metrics that provide insight on additional interaction patterns than traditional transaction log analysis.
CHAPTER 3

RESEARCH OBJECTIVES

The present research will attempt to contribute a comparative study of search behaviors across devices of different screen size to the existing literature. The iPhone and iPad were selected specifically for the study because of their majority share in the U.S. consumer smartphone and tablet markets (Table 8 and Table 9, respectively). Figure 8 demonstrates the screen size dimensions on all three devices.

Table 8 U.S. consumer smartphone market share (The NPD Group, 2014)

<table>
<thead>
<tr>
<th>Phone</th>
<th>Market Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPhone</td>
<td>45%</td>
</tr>
<tr>
<td>Samsung</td>
<td>26%</td>
</tr>
<tr>
<td>LG</td>
<td>8%</td>
</tr>
<tr>
<td>HTC</td>
<td>6%</td>
</tr>
<tr>
<td>Motorola</td>
<td>4%</td>
</tr>
<tr>
<td>(various others)</td>
<td>11%</td>
</tr>
</tbody>
</table>

Table 9 U.S. consumer tablet market share (Edwards, 2014)

<table>
<thead>
<tr>
<th>Device</th>
<th>Market Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPad</td>
<td>32.5%</td>
</tr>
<tr>
<td>Samsung</td>
<td>22.3%</td>
</tr>
<tr>
<td>ASUS</td>
<td>5%</td>
</tr>
<tr>
<td>Lenovo</td>
<td>4.1%</td>
</tr>
<tr>
<td>Amazon</td>
<td>1.9%</td>
</tr>
<tr>
<td>(various others)</td>
<td>34.2%</td>
</tr>
</tbody>
</table>

Using contemporaneously prominent devices allows an update to existing studies that use outdated devices (for example, Kamvar and Baluja (2006) studies comparing behaviors across conventional computers, PDAs, and traditional cell phones) and to further the body of work examining behaviors on mobile devices by focusing explicitly on the behavioral differences elicited by variations in screen size.
Research Objective 1. This study first seeks to compare information search efficiency and the quality of search outcomes on each device against both objective and subjective metrics.

[H1] The time spent on a search task is inversely proportional to the screen size of the device.

[H2] The probability of task success increases as screen size increases.

Aula, Khan, and Guan (2010) use these metrics of time on task and task outcome to understand the impact task difficulty as did Lorigo et al. (2006) in a study testing the influence of task type and gender on search behavior. Subjective measures were also implemented based on a general belief that smaller devices are more difficult to use.
[H3] The confidence of the participant during a search task on a given device is directly proportional to the screen size

[H4] The effort required by the participant to complete a task on a given device is inversely proportional to the device screen size.

Research Objective 2. The present study also uses a grounded theory approach to investigate how device size impacts the constitutive behaviors of search on small screen sizes. Grounded theory methodologies “allow the researcher to develop a theoretical account of the general features of a topic while simultaneously grounding the account in empirical observations or data” (Martin & Turner, 1986, p. 141). This approach is particularly useful in studies such as the present one, where existing theories describing behavior are sparse or abstract. The behaviors measured were both inspired by previous studies and experience using mobile devices. These behaviors are loosely categorized here as information browsing behaviors, query behaviors, and search navigation behaviors and will be discussed below:

Information Browsing

(1) Cumulative time on search result page – Time spent on the SERP (versus content screens, see below) was aggregated from each trial to describe the proportion of time the participant, spent interacting with the search engine. This includes browsing the result links and reading content preview, formulating and refining searches. This measure is intended to show the degree to which relevant information is gleaned from the search page as the number of entries and the quantity of information per entry varies by screen size. This was inspired by its use by Aula, Khan, & Guan (2010), who found that user spent more time on the search result page as the difficulty of the search task increased.
(2) *Cumulative time on content pages* – As a complement to measure (1), the time spent browsing, reading, and navigating through content on websites linked to from the SERP was measured and aggregated for each search session. As previous authors (i.e. Kamvar and Baluja (2007)) have used the amount of interaction occurring on a page as a proxy for the amount of information is gained from the page, this measure is intended to describe the degree to which relevant information is found on content pages.

(3) *Cumulative time scrolling* – To provide an additional dimension of behavior to measures (1) and (2), the cumulative time participants spent scrolling on both the search and content pages were measured independently. This metric is intended to describe the proportion of time participants spent browsing information as defined by Toms (2000): “an activity in which one gathers information while scanning an information space without an explicit purpose” (p. 423). This metric is also intended to exclude time when participants were formulating or typing queries, navigating through links, or fully reading content (this was captured by the following measure).

(4) *Cumulative time stationary* – In response to the previous measure, the cumulative time participants were stationary on both the search and content pages were measured independently.

(5) *Scrolling (number of results)* – Results (specifically, the title entries) viewed by participants were counted and aggregated for each session. Furthermore, each query leading to new results rankings were counted independently. This measure is intended to provide an alternative perspective to metrics describing how information is gleaned from the SERP.

(6) *Scrolling (number of screens)* – The number of screens of results was used as an
alternative description of scrolling behavior to account for the difference in number of entries on different devices: the 13” Macbook Air screen displays an average 6.5 search full result entries (i.e. title + url + abstract) per screen, the 9.7” iPad also displayed an average of 6.5 search results, and the 3.5” iPhone screen displayed an average of 2.5 results at the resolution used.

Search Queries

(7) Search query length – The total number of characters used in constructing the search queries for each task were counted as in Kamvar and Baluja (2007), Kamvar et al. (2009), and Aula, Khan, and Guan (2010), among others. This includes mistaken, deleted, or retyped characters, but not blank spaces. The purpose of this measure is twofold. First, it is intended to show the relationship between the length of the search bar and the construction of the textual search query. The second objective follows as a metric of complexity from Stojanovic, Studer, and Stojanovic (2003) assertion that queries may need to be more complex and specific to ensure the most relevant results are ranked highest on smaller devices.

(8) Number of query refinements – The number of query refinements was used to assess the process of revising the query to solicit the appropriate results in the space provided by each device (Shneiderman, Byrd, & Croft, 1997). A query refinement was counted as one that resulted in a new order of search results. This metric is intended to provide insight in how users rely on queries rather than browsing or navigating to view the appropriate result link.

Information Navigation
(9) Search branch length – This measure is based on White and Drucker’s (2007) concept of branch length and is used to understand whether the device screen size impacts the degree of multi-step website navigation by recording the number of “steps” away from the Google results page a user looks for information. For example, clicking a result link on the Google SERP leads to a branch length of 1. Clicking a link in the content of that page results in a branch with length 2, and so on.

(10) Rank of link clicked – The location of the links selected on the SERP (i.e. the first result is Link 1, the second result is Link 2, etc.) was recorded to describe how users perceived results in different locations on screens of different sizes. This measure was informed by Lorigo et al.’s (2006) study, which used it to compare the search behaviors of men and women.

(12) Click rate – Motivated by Kamvar and Baluja (2006) click rate was used to provide another perspective on the physical interactivity between the device and the user. This measure includes clicking on links to navigate (both on SERP and content screens), clicking to zoom-in to a section of the screen, and clicking to activate text box.
CHAPTER 4

METHODOLOGY

Participants

Thirty-six randomly selected Cornell University students (9 males, 27 females) participated in this study. Table 10 summarizes the participant demographics.

Table 10 Demographic summary

<table>
<thead>
<tr>
<th>Demographic Variables</th>
<th>Total Participant Group, N=36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>n = 27 (75%)</td>
</tr>
<tr>
<td>Male</td>
<td>n = 9 (25%)</td>
</tr>
<tr>
<td>Freshman</td>
<td>n = 7 (19.44%)</td>
</tr>
<tr>
<td>Sophomore</td>
<td>n = 7 (19.44%)</td>
</tr>
<tr>
<td>Junior</td>
<td>n = 6 (16.66%)</td>
</tr>
<tr>
<td>Senior</td>
<td>n = 11 (30.55%)</td>
</tr>
<tr>
<td>Graduate student</td>
<td>n = 5 (13.88%)</td>
</tr>
</tbody>
</table>

35 participants reported owning and using a laptop on a regular basis, with one participant regularly using a desktop. 32 participants reported owning and using an Internet enabled smartphone, 19 of which were iPhones. The remaining 4 participants owned conventional cell phones without Internet accessibility. Figure 9 demonstrates the composition of phone ownership by device. 9 participants reported owning and using an Internet enabled tablet device. Figure 10 depicts the ownership statistics of various computing and communication devices.
**Figure 9** Device ownership statistics.

**Figure 10** Phone ownership statistics.
Recruitment

Participants were recruited via flyers (Appendix C) placed in strategic locations across the Cornell University campus, and participants were screened to ensure prior familiarity with the Google Search tool and sufficient command of the English language to perform the required search tasks. Participants were given 10$ immediately following their session.

Experimental Design

A within-subjects Latin square experimental design was used, in which the order of both screen size and informational search task were randomized. Participants were instructed to complete two questions consecutively on each of three devices (designated as the randomized “block” variable.) The order of the question was also randomly assigned and designated as the “order” variable. Designated block and order variables allow for tests of such effects as learning or fatigue in the results.

Table 11 Representation of experiment design

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Device 1</th>
<th>Device 2</th>
<th>Device 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Block 2</td>
<td>Block 3</td>
<td></td>
</tr>
<tr>
<td>Order 1</td>
<td>Order 2</td>
<td>Order 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Order 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Order 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Order 2</td>
<td></td>
</tr>
</tbody>
</table>
Independent Variables

Research Objective 1.

[H1] Time on task – seconds.


[H3] Confidence – Participants rated perceived confidence using a Likert scale in which opposite sides were marked “very confident” and “very unconfident” [Appendix B].

[H4] Effort – Participants rated perceived effort using a Likert scale in which opposite sides were marked “little effort” and “a lot of effort” [Appendix B].

Research Objective 2.

Information Browsing

1) Cumulative time on search result page – seconds
2) Cumulative time on content pages - seconds
3) Cumulative time scrolling on search results pages - seconds
4) Cumulative time scrolling on content pages - seconds
5) Cumulative time stationary on search results pages – seconds
6) Cumulative time stationary on content pages - seconds
7) Scrolling - number of results
8) Scrolling - number of screens

Search Queries

9) Search query length – characters
10) Number of query refinements

Information Navigation

11) Search branch length – number of independent pages navigated to in succession
12) Rank of link clicked – link number clicked
13) Click rate – number of clicks
Equipment and Materials

Participants were instructed to complete six informational search tasks, two on each of three devices: a laptop, a tablet, and a smartphone. A standard 13” MacBook Air (325 mm x 227 mm, 1440 px x 900 px resolution, 128 ppi pixel density, iOS version 10.9.2) was used in the laptop condition and search tasks were conducted using the Google search site in a Firefox web browser. The Google search engine was selected due to its majority share in the U.S. explicit core search market (Table 1). Firefox was used specifically because it enabled the best control over search history. Search history needed to be re-set after each test on the laptop, within and between participants, to ensure that previously-clicked links were not highlighted (a common user interface feature in search engines), and that previously used search queries would not appear via the autocomplete feature (a standard Google search feature across all device platforms).

For the mobile conditions an Xcode emulator was used on the same MacBook Air laptop to simulate an iPad screen (241.2 mm x 185.7 mm, iOS version 7.1) and the iPhone screen (115.2 mm x 58.6 mm, iOS version 7.1). An emulator was used to control for input mechanism across devices. Table 9 depicts the conversion of multi-touch gestures to mouse gesture. Normally, these devices would have resolutions of 1024 px x 768 px resolution at 132 ppi pixel density and 960 px x 640 px resolution at 326 ppi pixel density respectively. Because a simulator was used, however, pixels were mapped one-to-one (and thus standardized) between the simulated device and the laptop at the laptop’s given pixel density of 128 ppi.

As with the browser in the laptop condition, the browser used in the emulator was reloaded between tasks and participants. The Google search engine was loaded from the
Firefox application at the maximum dimensions for each device in the upright position. Participants were instructed to use a standard mouse (Logitech M510) for navigation and the MacBook Air keyboard for text input.

Participants’ on-screen behaviors were recorded using Quicktime’s screen-capture feature, and paper surveys were used to collect contextual information before, during, and after the test.

**Procedure**

Participants were presented with an initial survey collecting demographic information. Validated questions were used that are representative of contemporary knowledge and variable difficulty, and whose answers require the participant to look past the first few results using Google (Lorigo, Pan, Hembrooke, Joachims, Granka, & Gay, 2006) [Appendix A]. Closed questions were specifically used to ensure a measure of success and were read aloud to the participant to discourage them from exclusively using words verbatim from the prompt. The participant’s on-screen behavior during each task was recorded until the task was completed.

**Table 12** Closed, informational questions presented to participants

| Q1 | Who discovered the first modern antibiotic? |
| Q2 | What actor starred as the main character in the original “Time Machine” movie? |
| Q3 | On what data could you vote in the 2012 presidential primary in the state of New York? |
| Q4 | Where is the tallest mountain in the state of New York located? |
| Q5 | A friend told you that Mr. Cornell used to live close to campus – between University and Stewart Aves. Does anyone live in his house now; if so, who? |
| Q6 | Who won the gold medal in the men’s singles luge event at the 2006 winter |
Questionnaires were administered to participants to supplement quantitative data with contextual information about their experience in each condition. Participants were asked to complete a short survey after each individual search task to solicit information about their familiarity with the topic and effort expended arriving at the answer. When all six tasks were completed, they were presented with a post-test survey that solicited contextual information by gauging their familiarity with online informational searches, their experience with mobile computing devices, and how they use the Internet in general.

Analysis

Research Objective 1. A continuous mixed model was constructed to assess the relationship between time on task and device size using time as the dependent variable and question, order, and device as fixed factors. A Bonferroni correction was applied to adjust for multiple comparisons.

A generalized linear model was used to examine the relationship between the binary correct/incorrect responses and device size.

Generalized estimating equations were also used to describe the subjective variables confidence and effort, however a Bernoulli distribution (logit) was used in place of a Poisson distribution. The Bernoulli distribution was implemented because it is better suited for the probability density function of random variables measured using a discrete scale. Results were reverse log transformed to return results - \( \text{logit} (p) = \log \left( \frac{p}{1-p} \right) \) - to proper proportions, where \( \text{logit} (p) \) is the returned value from SPSS.
Research Objective 2. Continuous mixed models were developed in the same fashion as for time on task using cumulative time on search result page and cumulative time on content pages as dependent variables.

Generalized estimating equations employing a Poisson probability distribution to describe the distribution of the dependent variables search query length, query refinements, search length per query, number of clicks, number of results, and number of pages.

Finally, search branch length and the rank of link clicked were both converted to probabilities based on the frequencies of each. This data was implemented in a generalized estimating equation using a Bernoulli distribution. Results were log transformed using \[ \log(p) = \log \left( \frac{p}{1-p} \right) \] to provide proper proportions.

The criterion for statistical significance was \( p \leq 0.05 \).
CHAPTER 5

RESULTS

In this section major findings are presented with respect to both the hypothesized outcomes and descriptive metrics. A mixed model analysis of variance indicated that block and order did not have not significant effects on the main dependent variables.

*Research Objective 1, [H1] – Rejected*

Participants spent a mean 77.65 seconds on tasks conducted on the laptop (SD=78.84s), 84.86 seconds on tasks conducted on the iPad (SD=76.89s), and 100.54 seconds on tasks conducted on the iPhone (SD=97.77).

![Figure 11 Mean Time on task by device](image-url)
The results of a mixed model for a continuous variable show that device does not have a main effect on time on task (p=0.100).

*Research Objective 1, [H2] - Rejected*

Overall, 61.2% of questions were answered correctly (reflecting the findings of Lorigo et al. (2006) who found a 60% success rate using the same test questions). The rate of correct answers varied very little between devices as 66.7% of questions were answered correctly using the laptop, 56.9% on the iPad, and 61.1% on the iPhone. Accordingly, a generalized linear model found that the interactions between device and response correctness were not statistically significant (p=0.255).

![Bar chart showing correct answers by device](image)

**Figure 12** Correct answers by device

*Research Objective 1, [H3] - Rejected*

**Confidence.** Self-reported confidence scores (between 0-not at all and 1-extremely) showed very little variability across devices. On average, participants rated their confidence using the laptop at 0.69 (SD=0.29), whereas the iPad received a confidence
score of 0.74 (SD=0.25), and the iPhone received a mean score of 0.75 (SD=0.24). Accordingly, device did not have a significant main effect on participants’ perceived confidence using the various devices to accomplish test tasks (p=0.139).

Figure 13 Self reported confidence score by device

Research Objective 1, [H4] - Rejected

*Effort.* Similar to confidence, self-reported scores of perceived effort (rated between 0-little effort and 1-a lot of effort) had relatively low variability between devices. Participants rated the effort required to complete tasks on the laptop at 0.37 (SD=0.24), 0.41 on the iPad (SD=0.24), and 0.45 on the iPhone (SD=0.29). Overall, device did not have a significant effect on effort rating (p=0.093).
Research Objective 2 - Information Browsing

Cumulative time on search and content pages. Mixed models of time on the search results page and time on content pages show a similar trend as that for total time on task.
Figure 15 Mean time on Google result page in seconds. Laptop SD=49.69, iPad SD=34.21, iPhone SD=58.97

Figure 16 Mean time on content page in seconds. Laptop SD=34.43, iPad SD=59.07, iPhone SD=47.33
The mean times on search and content pages can also be visualized as components of the total time on task (figure 17):

A test of model effects found no significant main effect of device on either time on the search page (p=0.270) or the content page (p=0.205). Further, a test of the proportion of time spent on the search result page compared with total time was also not statistically significant with respect to device as a fixed effect (p=0.378).

*Cumulative time scrolling and stationary.* The time spent on search and content pages was split into the time participants were stationary on the page and the time they spent scrolling, and results shows that time was spent differently between devices.
Participants were stationary on the search page for an average of 18.44 seconds in the laptop condition (SD=24.77s), 19.42 seconds on the iPad (SD=15.70s), and 15.57 seconds on the iPhone (SD=17.81s). Conversely, participants spent an average of 2.08 seconds (SD=5.37s) scrolling through the search results page on the laptop, 3.26 seconds scrolling on the iPad (SD=8.46s), and 11.5 seconds of scrolling on the iPhone (SD=19.01s). While the differences between the times stationary were not statistically significant (p=0.158), the differences between the amounts of time spent scrolling on each device were highly significant (p<0.001).

Figure 18 Mean time scrolling and stationary on search pages
Similar interactions are seen with the times spent scrolling and stationary on content pages as well. Participants spent an average of 25.85 seconds (SD=26.50s), 28.33 seconds (SD=44.87s), and 21.58 seconds (SD=22.60s) stationary on content pages viewed on the laptop, iPad, and iPhone respectively. A test of device effects found that these differences were not statistically significant (p=0.221). Participants spent an additional mean 10.17 seconds (SD=16.14s) scrolling on content pages on the laptop, 8.51 seconds scrolling on the iPad (SD=11.74), and 23.61 seconds scrolling on the iPhone (SD=26.53). As on the search results page, the differences between times spent

**Figure 19** Mean times scrolling and stationary on content pages
scrolling on content pages viewed across devices were statistically significant (p<0.001). A pairwise comparison shows that the difference in time scrolling on the laptop and iPad was not statistically significant (p=0.275), however there were significant differences between iPad and iPhone (P<0.001), and the iPhone and laptop (p<0.001).

**Number of search results viewed.** The number of results viewed by participants varied substantially by device. Participants viewed a mean 11.92 results on the laptop (SD=14.42), 10.11 results on the iPad (SD=8.66), and 6.88 results on the iPhone (SD=7.73). Accordingly, a test of model effects found a significant main effect of device on the number of result viewed (p=0.004).

![Graph showing number of Google search results viewed by device](image)

**Figure 20** Number of search results viewed

Furthermore, a pairwise comparison indicated statistically significant differences between the laptop and iPhone (p < 0.001), and the iPad and iPhone (p = 0.001) (though the difference between laptop and iPad was not statistically significant (p = 0.191)).
**Number of search result pages viewed.** The number of pages of results viewed per task, as the number of results presented on a single page varied by device, was analyzed. Participants viewed an average of 4.29 pages of results on the iPhone (SD=3.09), whereas they only viewed an average of 2.21 pages (SD=0.65) and 1.33 pages (SD=1.32) on the laptop and iPad, respectively. However, the effect of device was not statistically significant (p=0.116).

![Figure 21 Number of results pages viewed](image)

**Research Objective 2 - Search Queries**

**Query length.** A test of cumulative query lengths found very low variability in the number of characters used per query on each device. On average, participants used 44.47 characters for queries on the laptop (SD=40.93), 44.90 characters on the iPad
(SD=26.32), and 43.82 characters on the iPhone (SD=52.98). The effect of device effects on cumulative query length was not statistically insignificant (p=0.238).

![Cumulative search query length](image)

**Figure 22** Cumulative search query length

**Number of query refinements.** Queries were refined an average 0.72 times on the laptop (SD=1.64), 0.64 times on the iPad (SD=1.14), and 0.61 times on the iPhones (SD=1.49), with a range from 0 to 10 query refinements.
While it appears that participants made fewer query refinements as screen size decreased, a test of model effects with the number of refinements as the dependent variable indicated that the relationship was not statistically significant (p=0.823).

**Research Objective 2 - Information Navigation**

**Search branch length.** Participants navigated to the first step away from the SERP an average of 2.97 times using the iPhone (SD=0.07), 2.72 times using the iPad (SD=6.17), and 2.67 times using the laptop (SD=0.16). Participants navigated from the first to the second level an average of 0.42 times on the iPhone (SD=0.04), 0.33 times on the iPad (SD=0.09), and 0.03 times on the laptop (SD=0.12). Finally, participants navigated to the third level of website an average 0.08 times on the iPhone (SD=0.03), 0.11 times on the
iPad (SD=0.05), and 0.03 times on the laptop (SD=0.02). No participants navigated past the third level of link depth beyond the search result page.

![Figure 24](image.png)

**Figure 24** mean frequencies with which participants navigated to a branch step

The probability that participants navigated to a certain level in a search was calculated by comparing an individual’s frequency to the total times participants navigated to a given level, and applied to a generalized estimating equation. Screen size did not have a main effect on the probability of participants’ navigating to a specific branch step.
Table 13 Probability of a participant navigating to a website link level

<table>
<thead>
<tr>
<th>Level #</th>
<th>Mean probability of navigating to a website link level</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laptop</td>
<td>iPad</td>
</tr>
<tr>
<td>1</td>
<td>0.79</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>0.52</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Link rank clicked

Participants clicked on the first link presented on a search results page a mean 1.5 times on average per search session on the iPad, 1.06 times on the laptop and 1.03 times on the iPhone. Participants clicked on the second link 0.75 times on average per search session on the iPhone, 0.61 times on the laptop, and 0.58 times on the iPad. Figure 25 shows the decreasing frequency with which participants selected links ranked beyond the first two items. A generalized estimating equation constructed using the probability that participants clicked on a certain link (compared with all clicked links) across devices and found that there screen size had no main effect on the link rank clicked.
Figure 25 mean frequencies with which links were clicked.

Table 14 Probability of clicking a link on the results page.

<table>
<thead>
<tr>
<th>Link #</th>
<th>Mean probability of clicking link</th>
<th>P value.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laptop</td>
<td>iPad</td>
</tr>
<tr>
<td>1</td>
<td>0.60</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>0.09</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Click Rate. Participants clicked more on the iPhone than either the laptop or iPad with a mean rate of 4.75 clicks (SD=4.65) per search session compared with 4.21 clicks on the laptop (SD=3.63) and 4.28 clicks on the iPad (SD=4.35). Device did not, however, have a significant main effect on click rate (p=0.501).

<table>
<thead>
<tr>
<th>Device</th>
<th>Mean</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laptop</td>
<td>4.21</td>
<td>3.55</td>
<td>4.77</td>
</tr>
<tr>
<td>iPad</td>
<td>4.28</td>
<td>3.62</td>
<td>4.84</td>
</tr>
<tr>
<td>iPhone</td>
<td>4.75</td>
<td>4.05</td>
<td>5.48</td>
</tr>
</tbody>
</table>

**Figure 26** Number of clicks.
CHAPTER 6

DISCUSSION

The goal of this study was to test whether device screen size affected search engine usage. All four original hypotheses were rejected, indicating that screen size was not found to have a main effect on task efficiency or quality. Despite this finding, this study has raised compelling issues and questions that can inform the design of future studies, many of which have gone unarticulated in the absence of other comparative studies on search behavior.

A high degree of statistical variability was found across all metrics (despite log transformation and removing outliers from the datasets.) While this could be symptomatic of a small sample size, it also speaks to the assertions of Johnson and Meischke (1993) and Wilson (1981) on the role that individuals’ personal characteristics play in their search behaviors. This study used fairly high-level metrics of behavior that were not necessarily designed to capture the nuances of interpersonal variation in search strategies. This is particularly true in the case of query related behaviors, which were measured in terms of cumulative query length and the number of refinements made during a search session. Observing the content, complexity, and advanced operators (Aula, Khan, & Guan, How does search behavior change as search becomes more difficult?, 2010) of participants’ queries, the first step in information search, could provide rich insight to how people differ in setting up their search processes. Variation may also exist between individuals’ strategies for search result evaluation, as concluded by Aula, Majaranta and Raiha (2005) in an eye tracking study, however the technical limitations of eye tracking in the mobile context remain challenging to overcome. Understanding the degree of interpersonal variation in strategy could be a tremendously
valuable contribution to this space, as it would contextualize the individual findings on behaviors outlined in the literature review and described by this study.

The mean time on task found in this study was almost double the amount of time participants spent on the same set of questions in Lorigo et al.’s (2006) study, (77.65 seconds in the laptop condition of the present study compared with 46 seconds using the same sample size and comparable sample demographics.) The discrepancy in these times may be attributable to differences in the methodologies employed by these two studies: Lorigo et al. (2006) instituted a two minute time limit to all search tasks whereas the present study placed no restrictions on time.

The decision not to limit participants’ time was made to allow for more behavioral observations. In reality, however, users may not necessarily continue searching until they find the information they are looking for; they may stop prematurely when they perceive the cost of searching to exceed the benefit of finding the correct information. To be able to generalize behavioral studies to in vivo mobile usage, researchers need to understand where users’ internal threshold for search time is and what factors affect it. These factors can relate to both the topic of the information search and the ambient environment, and their impact does not end with the amount of time spent on a search. Are people conducting the same type of search on mobile devices as they are on conventional computers? How do the topics of search differ when the user is in motion and their attention is split between the device and the ambient environment, versus when they are stationary? How does this impact the amount of time they are willing to spend conducting search in these different contexts?
In a log analysis of over 1 million searches, Kamvar and Baluja (2006) found that, indeed, people were searching different topics on conventional computers compared with PDAs, and began to explore the relationship between topics and various high-level measurements of behavior. Search outcome was notably missing, however, because the authors were unable to ascertain whether users found the desired information before ending the search session. These questions are not only salient, but also difficult to quantify and control as an experiment given the breadth of variables to be monitored, and the technical limitations of measuring behavior on mobile devices.

Interestingly, despite the imposed time limit, Lorigo et al. (2006) also demonstrated similar task outcome rates (i.e. correct answers) as the present study. Approximately 60% of questions were answered correctly in Lorigo et al.’s (2006) study (conducted using a laptop) compared with 66.7% in the laptop condition, 56.9% on the iPad, and 61.1% on the iPhone in the present study. The increased time and consistent response may be explained by Kamvar and Baluja’s (2007) assertion that, given the opportunities and resources, people tend to do more exploration in a search session. It is more likely due to the fact that both studies (Lorigo et al.’s (2006) and the present study) used the same questions submitted to the same search engine. Despite having been conducted in 2006, the answers to the questions used in Lorigo et al.’s (2006) study may have been located on the same web pages as those where answers were found in the present study. These webpages may have had comparable link rank and search branch length, though neither study controlled for the location of the appropriate information.

Given the limited range of links clicked in the present study, certain questions emerge surrounding the impact of information placement (that is, where relevant results
appear on the search result page) on searching behaviors. To some extent these recall the questions voiced earlier in the Discussion on the types of searches people conduct on mobile versus conventional computing devices: are mobile users more likely to search for concrete, factual information whose location is easily accessible from a mobile device (as hypothesized by Aula, Khan, and Guan (2010)) and reserve more abstract search topics for the conventional search environment?

Despite the fact that the questions were designed to “ensure that the most intuitive queries would not always result in top-ranked results” (Lorigo, Pan, Hembrooke, Joachims, Granka, & Gay, 2006, p. 1125) the prompts were concise and well defined, and resulted in relatively efficient query formulation. The observed distribution of link ranks clicked in this study indicates that participants were able to formulate the queries sufficiently accurately such that relevant results were usually found among the first few links. They were not well suited to showing how participants would handle more open-ended search tasks in which the rankings of potentially relevant pages were distributed more broadly across the search page rather than the narrow distribution observed in this study. To take this one step further, controlling the for location of relevant information could extend the present findings from an empirical test of links that participants were prone to see, to a model of how participants respond to the full range of ranks on a search result page.

Further investigation into how participants spent their time showed that they were on search results pages for slightly longer than the participants of Lorigo et al.’s (2006) study. Participants in the present study spent approximately 41 seconds on the search page in the laptop condition compared with 26 seconds in Lorigo et al.’s (2006) study.
Adjusting for the overall differences in time on task between the two studies, participants in the present study spent a mean 53.5% of total time on Google search results pages in the laptop condition compared with 40.2% in Lorigo et al.’s study (2006). Participants in both studies spent close to equal amounts of time on search and content pages, contradicting the conclusions of Toms (2000), who asserts that “users make quick selection decisions but spend proportionally more time examining the content of articles” (p. 447). Toms’ study was designed to test different types of information environments that were somewhat more constrained than the naturalistic search information environment used in this study, which may explain the differences in attention given to the search results and content pages. Regardless, the behavioral measures implemented in the present study were better suited for the search results page than content pages. This study did not implement a descriptive metric of behavior on content pages outside of time spent scrolling and stationary, and search branch length.

The present study found that participants spent comparable amounts of time stationary on both search results and content pages in all three conditions. Participants spent significantly more time scrolling, however, on the iPhone than on either the laptop or the iPad during search sessions. These results are intuitive considering that far less information is viewable on a screen with reduced size, necessitating scrolling to view more content. Furthermore, this finding speaks to information foraging theory (Pirolli & Card, 1999) and users’ capacity to augment their search behaviors to maximize information processing. Despite scrolling more in the iPhone condition, participants viewed far fewer search results on the iPhone (5.97) compared with the iPad (8.79) or laptop (10.38).
It is likely that, despite having access to more information on large screens, users’ natural, but limited, F-pattern scan path (described in Lorigo et al. (2006)) expose them to fewer search results on the large screen than anticipated. This is supported by the finding that participants tended to select links of the same rank regardless of screen size. Specifically, this study found that participants clicked on the first link 60% of the time and the second link 24% of the time in the laptop condition. These probabilities are similar to those found in Lorigo et al. (2008) (in which participants clicked the first and second links 50.4% and 21.1% of the time, respectively) and do not vary substantially with the probabilities of clicking these links in the iPad and iPhone conditions (see Table 14).

Prior to the user selecting a link, however, they must browse the search results elicited by their query. Measurements of browsing tend to be somewhat challenging to operationalize given the technical limitations of measuring behavior on mobile devices, and in particular the lack of eye-tracking capabilities on these devices. Toms (2000) suggests, “browsing in fact may not be a quantifiable, defined task whose success is determined solely by its resolution” (p. 447). This study used the process measures of time stationary and scrolling to help describe browsing behaviors, however comparison across devices using these metrics is insufficient. For example, scrolling may be more indicative of browsing content on small devices, where users are constrained to scroll to view content, but is a poor metric on large screens which do not require users to scroll to view content. Eye tracking is generally used to distinguish the various activities users engage in on search pages that do not have physical manifestations (i.e. scanning, in-detail reading, etc.) In the absence of reliable eye tracking capabilities for the mobile
environment, it is very difficult to properly quantify this metric for comparison across conventional and mobile devices.

Screen size did not appear to have an effect on either query length or the number of query refinements, however the descriptive means for both differed from those found elsewhere in the literature. This study determined a mean query length of 44.47 terms and 0.72 refinements in search sessions on the laptop, 44.90 terms and 0.64 refinements on the iPad, and 43.82 terms and 0.61 refinements on the iPhone. Kamvar and Baluja (2006) found an average of 35.65 terms per search task, and Kamvar, et al. (2009) found an average of 18.72 terms per sessions in laptop search and 18.25 terms on iPhone search. Kamvar et al. (2009) also observed an average of 1.94 queries per search session conducted on the computer and 1.82 queries per search session conducted on the iPhone, whereas Kamvar and Baluja (2006) 1.6 queries per search session in a device agnostic study.

Screen size may not have had a substantial effect on query length or the number of query refinements because the mode of input was held constant. Participants entered queries using a standard MacBook keyboard and mouse in all three conditions, and were permitted to use both hands freely to type queries. This decision was necessary to isolate the visual effects of screen size from interaction mechanism. In reality, however, the direct manipulation, on-screen keyboards of these devices are much smaller than the physical keyboard used in this study and with conventional computers, and users may not necessarily have full use of both hands to type in the mobile context. Kamvar et al. (2009) assert that the size limitations of the on-screen keyboards on mobile devices make textual input and interaction more difficult than when using a physical keyboard. If this
is the case, it could result in different strategies for constructing and inputting search queries that may have implications beyond just query behavior.

Kamvar et al. (2009) point out that, due to the physical challenges of interacting with information in a small form factor environment, mobile users may rely on more focused queries that are refined more frequently. The authors hypothesize that users would employ various query formulation techniques to manipulate the ranks of relevant results to locations that are more accessible on a small screen (i.e. closer to the top of the screen), thus reducing the amount of scrolling required.

The consistency of the query metrics found in this study may also relate to the selection of search tasks used, and echoes the questions surrounding the types of queries that are made in highly mobile contexts (voiced previously in this section.) Mobile devices are used in a broad range of environments – spanning from highly active contexts in which users are in motion and their attention is divided between the device and the ambient environment, to the sedentary contexts in which conventional computers can be used. Though the participants of this study were required to answer specific questions, it remains unknown if these questions were representative of the types of search tasks individuals are likely to conduct on mobile devices in reality. The standardization of questions then provides insight to how individuals might approach any search task on a given screen, though it is not necessarily representative of how the device is actually used.
CHAPTER 7

CONCLUSION

The behaviors observed and reported in this thesis serve to provide a more thorough understanding of the ways in which mobile devices with reduced screen sizes impact information search processes. Data analysis resulted in the rejection of all four hypotheses, indicating that participants were equally proficient in information search on all three devices, in particular with respect to search efficiency and outcome. Furthermore, analysis of the various descriptive metrics of behavior demonstrates a number of differences in search behavior across screen size. In particular, the amount of time spent scrolling, and the number of results viewed, varied significantly across screens.

Taken as a whole, these results also raise and shape important questions that will aid future researchers in designing salient and generalizable studies. These issues can be broadly categorized as pertaining to the types of query topics relevant to the mobile context, the technical limitations of behavioral monitoring in the mobile context, and accounting for interpersonal variability in search strategy. Furthermore, the multitude of questions raised by this study highlights the fact that the results expressed here are highly dependent on the scope of this project. Information search is an inherently varied and complex space, and mobile search even more so. This study has attempted to apply a degree of experimental rigor to the space at the expense of particularizing the results. Despite this outcome, it has hopefully established a precedent for thinking critically about the impact of device form factor on information search processes.
CHAPTER 8

LIMITATIONS

The limitations of this study are generally attributable to the sample used and the technical limitations of collecting behavioral data on mobile devices.

In using exclusively undergraduate and graduate university students, the present study tested a fairly homogenous participant sample. Statistically, the participants belonged to a demographic that tends to have access to, and be more proficient with, alternative information search platforms and smaller screens (Duggan & Smith, 2013). Indeed, the Pew Internet & American Life Project (Duggan & Smith, 2013) shows that mobile Internet adoption is highly variable across demographics of age and educational attainment. It is possible that a sample more representative of the entire population would show that individuals who are less familiar with consuming information on small screens would yield very different results than the ones discussed here.

While comparable to other similar studies, the sample size of the present study was also relatively small with 36 participants yielding 216 search tasks. Analysis of the data demonstrated large statistical variability in the majority of behavioral measures, which can be symptomatic of insufficient sample size. Log transformation of the data and removing outliers, however, did not impact the significance of the behavioral measures, suggesting that the statistical variability found may be a product of a high degree of interpersonal variability rather than sample size.

Interpersonal variation occurs both within and between individuals and, as suggested by White and Drucker (2007), “…cognitive style, search task, background, information organizing strategies, and search experiences, to name but a few, are behind the variations in behavior” (p. 28). Despite this assertion, characterizations of these types
of variation remain fairly fractured in the literature, and thus were disregarded in the present study (as either a factor in participants’ behavior or as a factor in selecting participants with particular strategies.) Incorporating metrics of interpersonal variation would not only improve the generalizability of future studies, but aid in transforming observations of behavior into design tools.

This study also did not consider the full range of input mechanisms intended to compensate for the presumed challenges of using a small device. Recent iterations of smartphones and tablets have incorporated multimodal input (such as voice (e.g. Apple’s Siri), gesture (e.g. Graffiti), and image input) and alternative keyboards optimized for the mobile environment (e.g. 8pen, Android’s Swipe). It is possible that mobile users are relying on these alternative input mechanisms to support the optimal usage of their device. The present study is limited to the original forms of input mechanisms in an attempt to create a baseline understanding of search behaviors, however representing frequently used alternative functions would provide a salient perspective on the results.

As in any nascent field, technological challenges imposed substantial limitations on the present study. In particular, the lack of mobile sensing capabilities, and the poor state of mobile eye tracking had the largest impact on the design of this study.

The highly controlled nature of this experiment was partially based in a desire to be able to quantify certain behaviors with a high degree of confidence, and partially based on necessity. To conduct this experiment in a natural setting, participants would have needed devices with an exceptional multitude of sensors to account for all of the possible contextual variables of their search session. These include, but are certainly not limited to, the position of the device with respect to the participant, the characteristics of
the ambient environment, the physical state of the participant (i.e. motion), etc. Conducting tests in a controlled laboratory environment vastly diminished the need for additional measurements, at the expense of drastically reducing the generalizability of the results.

The limitations of eye tracking technology is included in this category of mobile sensing technologies, but its impact is so large on this study that it is worth calling out independently. Eye tracking technology emerged in the early 1990s, but was (and remains) largely constrained to stationary devices. Bulling and Gellersen (2010) attribute the lack of development in mobile eye tracking technologies to eye tracking accuracy, eye movements, calibration quality, calibration drift, and distinguishing between intentional gaze and accidental or coincidental gaze. The systems that exist are highly experimental and were not accessible for the purpose of this study.

This had the effect of limiting the granularity of the measures used to address information browsing behavior, and thus the accuracy of the results. As mentioned in the Discussion section, it became difficult to compare meaningful browsing behaviors across devices in the absence of eye tracking capabilities. As users are browsing search results, they may be completing a number of tasks, frequently simultaneously: reading, scanning, evaluating results, negotiating options, etc. Eye tracking studies can, and sometimes do, provide a nuanced view of how the search result page is used on conventional devices. Without the ability to compare this behavior with those on small screens, however, we can only speculate on the differences. Furthermore, because users are physically constrained to interact with information differently on different sized screens, the metrics that were used as proxies for certain behaviors (e.g. scrolling for scanning content)
became much less meaningful. The measures used in fact are only truly able to describe categories of behaviors. For example, the amount of time spent on search result pages and content pages are only able to describe the cumulative amount of time participants were engaging in countless other more nuanced actions.

This study attempted to provide insight on mobile information search within the constraints of technology. As the tools emerge to address more nuanced behaviors the metrics used here should be updated and reformulated to reflect the current state of technology, however the present research provides a reasonable baseline for understanding search behaviors on small screens and inspiring future studies on alternative devices.
APPENDIX A

QUESTIONS & ANSWERS

[Q1] Who discovered the first modern antibiotic?
[Q2] What actor starred as the main character in the original “Time Machine” movie?
[Q3] On what data could you vote in the 2012 presidential primary in the state of New York?
[Q4] Where is the tallest mountain in the state of New York located?
[Q5] A friend told you that Mr. Cornell used to live close to campus – between University and Stewart Aves. Does anyone live in his house now; if so, who?
[Q6] Who won the gold medal in the men’s singles luge event at the 2006 winter Olympics in Turin?

[A2] Rod Taylor
[A4] Keene, Essex County, NY
[A5] Yes, Delta Phi Fraternity
[A6] Armin Zöggeler
APPENDIX B
SURVEYS

1. Pre - Test Survey

**Participant ID:**

**Pre-Test Survey**

1. You are:
   - O Female
   - O Male

2. You are a(n):
   - O Freshman
   - O Sophomore
   - O Junior
   - O Senior
   - O Grad student

2. Post – Task Survey (x6 Tasks)

**Participant ID:**

**Task (after each task is completed)**

1. Answer:
   
   Please mark with an “X” along the axis.

2. How confident are you in the answer you found?

   Very unconfident

   |__________________________________________________________|

   Very confident

4. How much did you know about the subject matter of the question before this task?
Nothing at all

3. How much effort did your search require before you found the answer?

A lot less than expected

A lot more than expected

3. Post – Test Survey

**Participant ID:**

**Post-test Survey (after all tasks are completed)**

1. How comfortable do you feel using the Internet?

Very uncomfortable

Very comfortable

2. How satisfied are you with your current skills for using the Internet?

**Very unsatisfied -
I can’t do most things I want to do**

**Very satisfied -
I can do everything I want to do**

3. How comfortable do you feel using computers, in general?

Very uncomfortable

Very comfortable

4. How comfortable do you feel using your Internet-enable mobile devices, in general?

☐ N/A

Very uncomfortable

Very comfortable
5. Please indicate the electronic equipment you currently own and use: (check all that apply)
   - Desktop
   - Laptop
   - Mobile phone (NOT Internet enabled)
   - Smartphone
   - Internet enabled tablet
   - e-book device (IE Nook, Kindle, etc.)
   - MP3 player
   - None of the above
   - Other – please specify: ________________________________

6. Which of the following Internet – enabled mobile devices do you use? (check all that apply)
   - iPhone
   - iPod touch
   - iPad
   - Blackberry – specify model: ______________________________
   - Palm Pre – specify model: ______________________________
   - Droid – specify model: ______________________________
   - Nexus One – specify model: ______________________________
   - I don’t have one
   - Other – please specify: _______________________________

7. Approximately how many hours a day do you spend using Internet-enabled applications on your mobile device (ie. Not including voice, text) per day?

8. Approximately how many hours a day do you spend using Internet-enabled applications on your computer per day?

9. If applicable, to what degree do you use your Internet-enabled mobile devices for the following activities?
   - N/A
   a. Social networking
      Never ________________________________ Always
   b. Reading content (e-books, articles, etc.)
      Never ________________________________ Always
<table>
<thead>
<tr>
<th>Activity</th>
<th>Never</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>c. Getting news alerts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Accessing email</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Text messaging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Searching for information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Getting directions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h. Uploading content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Playing games</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j. Listening to music or watching videos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k. Completing coursework or participating in lectures</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. If applicable, *to what degree* do you use your **Internet-enabled computer** for the following activities?

☐ N/A
a. Social networking
Never  Always

b. Reading content (e-books, articles, etc.)
Never  Always

c. Getting news alerts
Never  Always

d. Accessing email
Never  Always

e. Text messaging
Never  Always

f. Searching for information
Never  Always

g. Getting directions
Never  Always

h. Uploading content
Never  Always

i. Playing games
Never  Always

j. Listening to music or watching videos
Never  Always

k. Completing coursework or participating in lectures
Never  Always
11. When searching for information, how frequently do you use the following search engines **on your computer**?

a. Google
   - Never
   - Always

b. Bing
   - Never
   - Always

c. Yahoo
   - Never
   - Always

d. AOL
   - Never
   - Always

e. Other:
   - Never
   - Always

12. When you search for information online, how frequently do you use the following search engines (please consider both full site and mobile application) **on your mobile device**?

a. Google
   - Never
   - Always

b. Google MOBILE
   - Never
   - Always

c. Bing
   - Never
   - Always

d. Bing MOBILE
   - Never
   - Always

e. Yahoo
<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>f. Yahoo MOBILE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. AOL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. AOL MOBILE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Other:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

STUDY RECRUITMENT MATERIALS

Are you a Cornell student, do you use Google, and want $10?

Contact mb2294@cornell.edu to participate in a short study!
APPENDIX D

INFORMATION SEARCH MODELS

Table 1 Ellis (1989a, p.178; Ellis, Cox, & Hall, 1993) Behavioral Model of Information System Design - Definitions

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting</td>
<td>Activities characteristic of the initial search for information</td>
</tr>
<tr>
<td>Chaining</td>
<td>Following chains of citations or other forms of referential connection</td>
</tr>
<tr>
<td></td>
<td>between material</td>
</tr>
<tr>
<td>Browsing</td>
<td>Semi-directed in an area of potential interest</td>
</tr>
<tr>
<td>Differentiating</td>
<td>Using differences between sources as filters on the nature and</td>
</tr>
<tr>
<td></td>
<td>quality of the material examined</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Maintaining awareness of developments in a field through the monitoring</td>
</tr>
<tr>
<td></td>
<td>of particular sources</td>
</tr>
<tr>
<td>Extraction</td>
<td>Systematically working through a particular source to locate material</td>
</tr>
<tr>
<td></td>
<td>of interest</td>
</tr>
<tr>
<td>Verifying</td>
<td>Activities associated with checking the accuracy of information</td>
</tr>
<tr>
<td>Ending</td>
<td>Activities characteristic of information seeking at the end of a topic</td>
</tr>
<tr>
<td></td>
<td>project, for example, during the preparation of papers for publication</td>
</tr>
</tbody>
</table>

Table 2 Shneiderman, Byrd, & Croft (1997, p. 3) Four Phase Framework for Search - Definitions

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation</td>
<td>What happens before the user starts a search</td>
</tr>
<tr>
<td>Action</td>
<td>Starting the search</td>
</tr>
<tr>
<td>Review of results</td>
<td>What the user sees resulting from the search</td>
</tr>
</tbody>
</table>
Refinement | What happens after review of results and before the user goes back to formulation with the same information need

### Table 3 Kuhlthau (1991, p. 366-368) Information Searching Process - Definitions

<table>
<thead>
<tr>
<th>Stage</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiation</td>
<td>Recognizing a need for information</td>
</tr>
<tr>
<td>Selection</td>
<td>Identify and select the general topic to be investigated or the approach to be pursued</td>
</tr>
<tr>
<td>Exploration</td>
<td>Investigate information on the general topic in order to extend personal understanding</td>
</tr>
<tr>
<td>Formulation</td>
<td>Form a focus from the information encountered</td>
</tr>
<tr>
<td>Collection</td>
<td>Gather information related to the focused topic</td>
</tr>
<tr>
<td>Presentation</td>
<td>Complete the search and to prepare to present or otherwise use the findings</td>
</tr>
</tbody>
</table>

### Table 2 Johnson and Meischke (1993, p. 346-350) Information Seeking Model - Definitions

<table>
<thead>
<tr>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Demographic factors</td>
</tr>
<tr>
<td>Direct Experience</td>
<td>An individual’s degree of direct experience with [the subject matter]</td>
</tr>
<tr>
<td>Salience</td>
<td>The perceived applicability of information to a problem that he or she has</td>
</tr>
<tr>
<td>Beliefs</td>
<td>Individual’s belief in the efficacy of [information]</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Message content attributes and communication potential</td>
</tr>
<tr>
<td>Utility</td>
<td>Relates the information provided by the medium directly to the needs of the individual</td>
</tr>
<tr>
<td>Actions</td>
<td>Actions that enable the purposive acquisition of information from selected information carriers, with three dimensions: method, scope, and depth</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


