Territories and Border Crossings:
Information Search and Location in the Online Environment

By

Brian Kimball Dunn

B.A., Brigham Young University, 1996

M.B.A., University of California, Irvine, 2002

Submitted to the Graduate Faculty of the

Joseph M. Katz Graduate School of Business

in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

JOSEPH M. KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Brian Kimball Dunn

It was defended on

June 30, 2014

and approved by

Dr. Kevin H. Kim, Associate Professor of Research Methodology, School of Education and Associate Professor of Business Administration, Joseph M. Katz Graduate School of Business, University of Pittsburgh

Dr. Laurie J. Kirsch, Vice Provost for Faculty Development and Professor of Business Administration Information Systems and Technology Management, Joseph M. Katz Graduate School of Business, University of Pittsburgh

Dr. Paul Benjamin Lowry, Professor of Information Systems, Information Systems Department, City University of Hong Kong

Dr. Narayan Ramasubbu, Assistant Professor of Business Administration, Joseph M. Katz Graduate School of Business, University of Pittsburgh

Dissertation Advisor: Dr. Dennis F. Galletta, Professor of Business Administration, Doctoral Program Director, Katz Graduate School of Business, University of Pittsburgh

Territories and Border Crossings:
Information Search and Location in the Online Environment

Brian Kimball Dunn, PhD

University of Pittsburgh, 2014

Table of Contents

List of Tables

List of Figures

Preface

This dissertation was based on three different data collections, two experimental (Study 1 and Study 3) and one archival (Study 2). Studies 1 and 3 are wholly unpublished work; a research-in-progress version of Study 2, which included work completed through the data coding process but prior to analysis, was presented at the poster session of the International Conference on Information Systems in 2012 in Orlando, Florida.

For the completion of this dissertation, I owe several debts of gratitude. First, my thanks to my dissertation committee members, Dr. Dennis F. Galletta (chair), Dr. Kevin H. Kim, Dr. Laurie J. Kirsch, Dr. Paul B. Lowry, and Dr. Narayan Ramasubbu. Their thoughtful comments, encouragement, and flexibility have ensured a constructive, positive, and efficient process.

Study 1 has benefitted from the assistance of Dr. Ramasubbu in the framing, story-identification, and analysis, Dr. Galletta, who expended considerable hands-on energy in collection of data and who was instrumental in assisting with the experimental design of the project, and Dr. Lowry, who offered timely and critical advice, particularly in identifying appropriate background literature and measurement instruments. I am also grateful for comments and suggestions offered by presentation attendees at the University of Pittsburgh and the University of Oklahoma.

Study 2 would not have come about without generous help and capable effort from Drs. Il Im and Dong-Il (Don) Lee, both of Yonsei University in Korea, as well as Dr. Galletta. Drs. Im and Lee obtained the unique and detail-rich archival data set, provided access to student coders, were pivotal in the analysis of the data, and helped keep the project energized and moving forward. Dr. Galletta offered important insight into the development of the hypotheses, coding scheme, and framework of the study, while keeping the process constructive and balanced.

Study 3 has profited immensely from the assistance of Drs. Ramasubbu and Galletta, who were instrumental in all stages of the research process. The study further benefitted from the workshop environment of the International Conference on Information Systems Doctoral Consortium 2013 in Milan, Italy, my attendance of which was made possible by the generous David Berg Center for Ethics and Leadership BNY Mellon Corporate Social Responsibility Doctoral Student grant. My gratitude also to Dr. Lowry for his thorough comments on an early draft, Dr. Alexandra Durcikova and numerous Amazon Mechanical Turk workers for their guidance and advice that facilitated data collection through a new-to-me medium, Jon C. Jensen for his timely coding advice, and to Steve Freitas for his generous donation of server space and network admin talents.

In addition, my thanks to Dr. Chris F. Kemerer for last-minute help with the initial draft of the dissertation proposal as well as for providing publishing opportunities, funding opportunities, and insightful strategic advice over the last four years; to Carrie Woods and Chris Fedor at the Katz PhD office for their selfless guidance and support; to the other professors at both Pitt and Carnegie Mellon University who offered course instruction and contributed substantially to the academic acculturation process; and to my PhD student colleagues, who were a frequent and necessary source of empathy and validation during this ordeal.

Finally, thanks to family members and friends, who provided a needed conduit to the real world and have not seemed to mind so much allowing me to pay less than my fair share at dinner over the last four years. Don’t worry, folks: I’ll be rich in three months.

Introduction

In the 23 years since the World Wide Web was released to the public, much has changed. No longer the sole province of universities and academia, Web use has exploded; the United Nations International Telecommunications Union estimates that there will have been over 2.9 billion Internet users worldwide in 2014 (2014). This growth has been accompanied, and likely in part driven by, advances in technology that include the wide availability of broadband, wireless networking, and improved display and server technologies.

As Internet usage has increased, so too has the Web's importance for businesses and other types of organizations. Websites are now ubiquitous. Firms, non-profits, schools, churches, and even Little League baseball teams are expected to own and maintain websites — in many cases, several of them. Firms can be expected to create unique websites based on various lines of business, sites need to be updated as prevailing browsing technology changes (e.g., new browser features or, historically, optimizing for wider screen widths), and companies are expected to adapt their websites to address the requirements of the increasing array of devices used to access the Internet.

Websites consume a significant portion of corporate budgets; a 2013 Gartner study suggested that 37.2% of surveyed firms' digital marketing budgets went toward website creation, maintenance, and analysis activities (Gartner, 2013). The costly scale of managing these sites has been borne out by the US government's recent HealthCare.gov website, which is believed to have cost at least $310 million to build and launch (Kessler, 2013). Beyond initial launch, website owners must also keep content up-to-date, maintain an infrastructure, and optimize performance based on analyses of website-generated metrics. Further, website owners spend considerable effort to attract visitors to their sites as online marketing and advertising activity cost companies tens of billions of dollars annually (Lunden, 2013).

Given this level of commitment, then, it is important that website owners get the most out of their websites. This can be difficult however, as the actual value of having a website is dependent on the actions of users, who may arrive at a website for a variety of purposes, who may be coming as part of a broader task, or who may not be consciously aware of which sites are being visited. This is an issue of particular importance when a website is merely one stop for a user during a broader online experience consisting of many sites and online interactions. Given these complications, how should website owners proceed?

Fortunately, academic literature has investigated a number of phenomena related to the behavior of users on websites. Among these phenomena are included flow (Koufaris, 2002), cognitive absorption (Agarwal & Karahanna, 2000), telepresence (Coyle & Thorson, 2001), interactivity (Albert, Goes, & Gupta, 2004; Coyle & Thorson, 2001; Lowry, Vance, Moody, & Beckman, 2008; Palmer, 2002; Song & Zinkhan, 2008), usability (Agarwal & Venkatesh, 2002; Lowry et al., 2008; Venkatesh & Ramesh, 2006), vividness (Coyle & Thorson, 2001), delay (Galletta, Henry, McCoy, & Polak, 2006; Palmer, 2002), annoyance (McCoy, Everard, Polak, & Galletta, 2007), trust and distrust (Gefen, Karahanna, & Straub, 2003b; Lowry et al., 2008; McKnight, Choudhury, & Kacmar, 2002; Pavlou & Fygenson, 2006), privacy (Hui, Teo, & Lee, 2007), design (Zhenhui Jiang & Benbasat, 2007; Nadkarni & Gupta, 2007), and user acceptance (Gefen et al., 2003b; Pavlou & Fygenson, 2006). This literature has made a great contribution to the field and provided numerous insights to practice.

We note, however, that the established literature leaves gaps by generally making one of two important assumptions. First, users have often been treated as having homogeneous intent or, said otherwise, studies have not sufficiently controlled for the different possible reasons that a user might have for visiting a given website. Instead, research has separately investigated behavioral effects pertaining to specific task-oriented intents of users (e.g., purchases from an e-commerce site) or hedonically motivated user intentions (e.g., reading about sporting events), but generally not examining both types of intentions within the same context to understand the differences.

Second, research has focused either on a single site in isolation or on broader user experiences. While it is certainly possible that a user goes online with the idea of visiting one and only one specific website, it seems reasonable that there are also situations in which users visit a broad set of websites within the same online experience. In this case, the quality of the user's experience at Site B may be influenced by the same user's experience at Site A that preceded the visit to Site B and at Site C directly followed it. The existing literature has generally left out the effects on a single site when located within a broader online experience.

To address these gaps, this dissertation invokes the metaphor of location and travel to describe users' actions on the Web. The location and travel metaphor has long been a part of the Internet context; a number of terms in the common parlance of online activity already imply these conceptualizations. A Web "site" inherently suggests the existence of a (metaphorical) location as do the concepts of a Web "address" and a uniform resource "locator" (URL). Further, people are said to "go to" a website and "follow" a link, both suggesting travel. Indeed, Web metrics are often described in terms of "visits" and "visitors". As in the real world, online "travelers" can move from location to location with different intents and, further, this travel among locations may impact the user's experience and thereby the effectiveness of the website investment.

We approach these topics over the course of three discrete studies. These studies progress from the isolated site outward as follows.

In the first study, we look at users' movements within individual websites (see Figure 1.1) to identify the significant properties and qualities of their within-site travel behavior that should be preferred by website owners. Further, we examine how that behavior can be affected by design decisions and how user intent impacts preferred behavior. In this study, we uncover an apparent conflict between the more IS-centric information foraging theory (Pirolli, 2007) and marketing-related theories of brand exposure (e.g., M. C. Campbell & Keller, 2003; Zajonc, 1968). These theories proffer seemingly contradictory recommendations for determining the ideal website visit. Information foraging theory emphasizes efficiency (less time on site), while brand exposure emphasizes more repetition of messages (more time on site). We reconcile these two theories by adding an additional component to information foraging related to the within-website locations traversed by a user, namely *territory*. We then conduct a laboratory experiment and analyze the resulting data to test hypothesized relationships between user intent (e.g., Browne, Pitts, & Wetherbe, 2007; D. J. Campbell, 1988; Simon, 1960), design characteristics of cognitive absorption (Agarwal & Karahanna, 2000; Lowry, Gaskin, Twyman, Hammer, & Roberts, 2013) and interactivity (Yuping Liu & Shrum, 2002), foraging time and territory, and brand engagement (Keller, 2008; Mollen & Wilson, 2010). We find that the user's intent *does* affect behavior and, further, that our reconciliation of the previously contradictory-seeming perspectives of information foraging and brand repetition is valid.

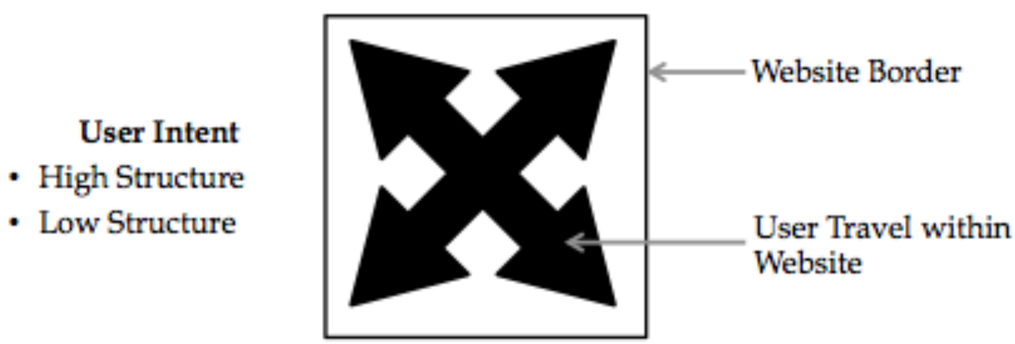


Figure 1.1. Study 1: Travel within a website.

Having seen that user intent has a significant effect on a user's behavior and that user behavior affects the potential for website success, we next address the need to better infer user intent based on how the user arrives on a website (see Figure 1.2). We address this question in the context of sponsored search advertising (e.g., Google AdWords) by using a data set of nearly 2.4 million sponsored search ad clicks collected by a large online travel agency in Asia. Similar to information foraging, we discuss the concept of information search and how it relates to user intent, identifying similarities between Simon's perspective on problem solving (1960), the marketing-based concept of the purchase funnel (e.g., Kotler, 1997; Lavidge & Steiner, 1961; E. S. E. Lewis, 1903), and the more Web-centric perspective of task structure (e.g., Browne et al., 2007; D. J. Campbell, 1988). Our analysis supports our assertion that user intent can be inferred based on the user's means of arrival onto a website.

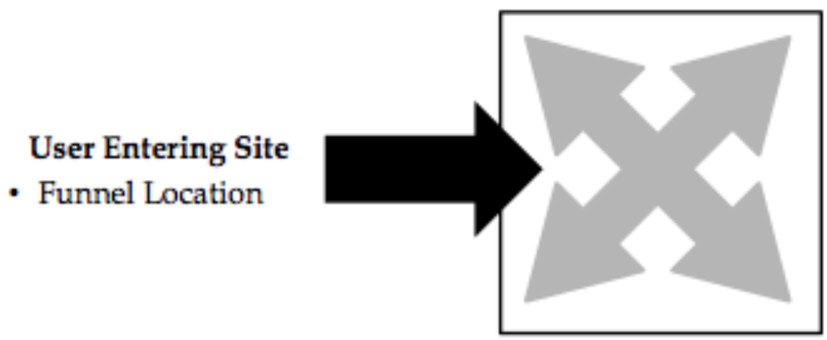


Figure 1.2. Study 2: Means of arrival on a website.

Finally, our third study builds on the first two by examining the effects of user travel *among* websites (see Figure 1.3). It has been reported by major websites that more users than ever are accessing their sites without viewing the site's homepage (Tanzer, 2014; Thompson, 2014). Even knowing which behaviors should be encouraged and understanding the effects of user intent (i.e., Study 1) *and* being able to better infer that user intent based on means of arrival to the site (i.e., Study 2), website owners can only reap the long-term benefits of a user's site visit(s) if that user recognizes having visited the site and is willing to attribute the resulting success of an experience to it. But when do users recognize which sites they have visited and how do they attribute credit to sites that help them complete their tasks? These questions directly reflect our earlier-stated concern about understanding the effects of a website existing within a broader online experience that includes many different sites.

To answer them, we introduce the concept of the website border, or the point at which a user crosses from one online location to another, whether perceptually or formally. We discuss the theory of space and place (e.g., Buttimer, 1976; Relph, 1976; Tuan, 1977, 2001), sourced from the field of experiential geography, which we relate to users coming to attach meaning to websites. In an experiment, we then manipulate both border strength and user intent and analyze our data to find that higher borders are helpful in encouraging users to recognize having interacted with a site as well as attributing credit to that site. We also find that the user's intent when visiting can affect the user's recognition and attribution.

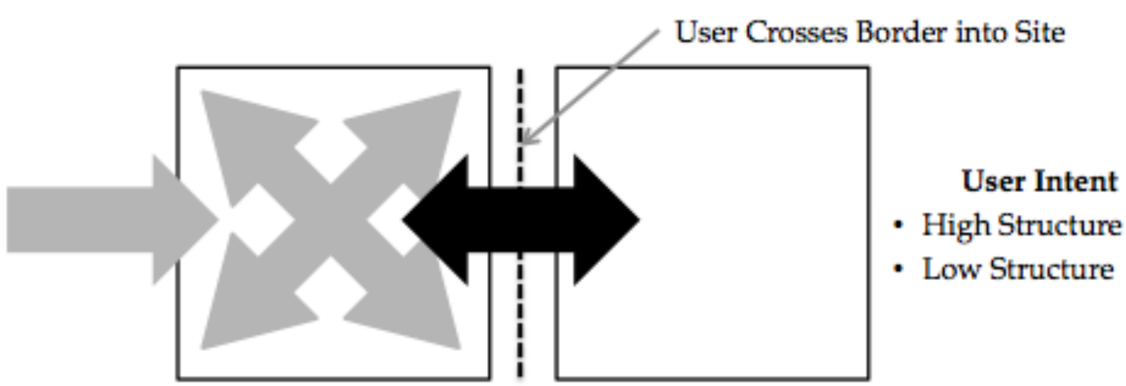


Figure 1.3. Study 3: Effects of traveling across website borders.

Taken together, the three studies of this dissertation make a substantial contribution both to research and practice. For research, we address key gaps in the literature, in particular analyzing the importance of user intent and a site's existence within a broader online experience. This should enable researchers to reconcile the concerns and predictions of different theories of website user behaviors while continuing to explore the evolving website interactions and online experiences in modern-day technological settings. For practice, the studies identify key problems and challenges faced by website owners. In particular, we increase understanding of the effects of user intents, site design properties, attracting and handling traffic, and enacting means to ensure that users both recognize having visited their sites and attribute credit for the experience facilitated by the sites.

Time and Territory: Preferred User Behaviors and Brand Engagement from the Website Owner's Perspective

2.1. Introduction

Given the considerable commitment website owners make to creating and managing their online presences (see Chapter 1), it is important that website owners keep in mind their sites' end goals and endeavor to design and maintain websites that achieve those goals. Reasonably, different websites have different objectives. Past information systems literature has primarily considered the objectives of e-commerce sites (Benbasat, 2006; Brynjolfsson & Smith, 2000; D. E. Campbell & Wells, 2013; Gefen, 2002; Z. Jiang & Benbasat, 2007; Lowry et al., 2008; McKnight et al., 2002; McKnight & Choudhury, 2006; Pavlou & Fygenson, 2006) and of sites seeking user loyalty and intent to use (Li, Browne, & Wetherbe, 2006; Mithas, Ramasubbu, Krishnan, & Fornel, 2007; Webster & Ahuja, 2006). In addition, many websites are intended to build a sense of brand engagement (Keller, 2008) between the website owner and its users. In this study, we take a novel look at the last of these contexts, that of the brand-building website, and focus on the issue of optimal user behavior in this context.

Existing literature on understanding optimal user behavior suggests two apparently contradictory approaches. In the marketing and advertising literature, conditioning through repeated brand exposure has been found to help brands associate themselves with positive concepts, which increase the probability of consumers taking actions favorable to the brand (Axelrod, 1968; Baker, Hutchinson, & Moore, 1986; M. C. Campbell & Keller, 2003). From this perspective, more-involved user website experiences should give the brand better opportunities for making a significant impression. Therefore, the longer a user spends on a website the better: with more time a user becomes more exposed to the brand and the ideas that surround it

On the other hand, information foraging theory (Pirolli & Card, 1999; Pirolli, 2007) suggests that a time-efficient experience is optimal, i.e., that the more quickly the user is able to complete the intended task (i.e., gather information, complete a transaction), the better. This perspective is supported by the practitioner design literature (Krug, 2009; Nielsen, 2000), where an efficient online experience is equated to a good brand experience. In this paper, we reconcile these two perspectives by suggesting that foraging behavior be measured not only through time spent in looking for information, but also in the territory covered during foraging behavior.

We also consider the actions that website designers can take to influence these two dimensions of foraging. Site designers and user experience specialists have created online experiences that have differing impacts on users. Some of these impacts have been measured in terms of interactivity, the degree to which a person and technology communicate (Huang, 2003; Yuping Liu & Shrum, 2002), and cognitive absorption, in which a user becomes lost in an information technology-mediated experience (Agarwal & Karahanna, 2000; Lowry et al., 2013). This effort is complicated by the fact that users may arrive at a website with different intentions and objectives in mind. We propose a set of hypotheses by which we predict how site design properties and a user's intent affect the user's foraging behavior both in terms of time and territory.

Finally, we introduce into the IS field the concept of brand engagement, which reflects both an attachment to a brand as well as an intent to act on that attachment (Keller, 2008). We then consider the impact of foraging on brand engagement, again with the notion of understanding what constitutes optimal user behavior from the website owner's perspective. These hypotheses are then tested using data collected in a laboratory experiment. Implications of our analysis are discussed.

2.2. Theoretical Background

2.2.1. Online Behavior: Brand Repetition and Information Foraging

2.2.1.1. Brand Conditioning and Repetition

Brand conditioning refers to the process by which a consumer comes to associate feelings and ideas with a brand (e.g., M. C. Campbell & Keller, 2003). Creating and strengthening this association through repeated exposure to images and ideas (e.g., through advertisements or online interactions on websites) dates back to Aristotle's rhetoric on habit formation[1] and Pavlov's conditioning experiments involving bell-ringing and salivating dogs (1927). Ultimately, the more frequently a consumer is

---

[1] From Aristotle's *Nicomachean Ethics.*

conditioned to an idea such as an emotional association with a brand, the stronger this idea becomes within the mind of the consumer (Axelrod, 1968; Baker et al., 1986; M. C. Campbell & Keller, 2003; Haley & Case, 1979; Pechman & Stewart, 1988).

The concept of branding through repetition, experience, and exposure has formed the backbone of advertising and brand messaging. It provides the rationale for companies that present themselves with consistent messaging, whether through in-store displays, television advertising, press releases, or website content and appearance. From the perspective of repetition and exposure, the more a user experiences a website, the greater the opportunity to strengthen positive brand associations (Baker et al., 1986; Harrison, 1977; Zajonc, 1968). With this in mind, many practitioners have used metrics such as pages viewed and time on site as measures of a successful user site visit.

2.2.1.2. Information Foraging

On the other hand, in more recent years the concept of information foraging has been introduced to the field of information systems. *Information foraging* refers to the process by which a system user seeks information, makes sense of it, evaluates it, and eventually stops to enjoy the fruits of that labor (Pirolli, 2007). The concept is derived from the ecological idea of foraging animals. Consider an herbivorous bird seeking food in meadow. The bird wants to maximize its intake of nourishment, while minimizing the energy expended to find it. In doing so, the bird may fly from shrub to shrub and hop from branch to branch seeking berries. Once it finds its quarry, the bird will eat the berries until sated or until food gathering at that location stops being cost-effective in terms of nourishment per unit of effort. This may be the case when all reachable berries are eaten, other birds have become aware of the food source, or predators have entered the vicinity. In selecting a location to begin feeding, animals will satisfice; in other words, they will not necessarily attempt to locate the absolute best source of food, but, rather, will identify the source that presents the best nourishment-to-effort ratio.

Similar patterns have been found among humans seeking information within the information systems context (Browne et al., 2007; Galletta et al., 2006; J. Liu, Zhang, & Yang, 2004; Pirolli & Card, 1999; Pirolli, 2007; Simon, 1955). Within the online environment, users will search a website trying to find their desired information. Once a promising avenue is found, users will follow this lead and then consume the information available there until the effort required to gain another unit of information exceeds the value of that information (Guo, 2001; Stigler, 1961).

We note that information foraging theory has a short history in aiding researchers to investigate phenomena in the information systems field, including user path modeling (J. Liu et al., 2004; Montgomery, Li, & Srinivasan, 2004), the effects of website response delay (Galletta et al., 2006), stopping rules for online search (Browne et al., 2007), choice of on-site information finding strategies (Katz & Byrne, 2003), and measurement of website navigability (Fang, Hu, Chau, Hu, & Yang, 2012).

In applying this idea of foraging to human behavior, Pirolli refers to Miller's assertion that humans are "informavores" and thus driven to acquire and consume information (1983). Along this line of thinking, it has been suggested that the need for information arose through evolution as a way of protecting oneself from a dangerous environment, much as prairie dogs stand atop their holes surveying the landscape for imminent threat (Dennett, 1991). Just as a foraging bird will find a patch of berries from which to eat, information foragers will likewise encounter "information patches", where they will stop and consume information. Within the Internet context, these patches are found within the "territory" of websites that contain information that aids the user in completing whatever task originally compelled the website visit.

Pirolli and Card construe information foraging as a series of information gathering and sense-making activities (1999). These activities, however, cannot be limitless in scope, especially given the almost infinite quantity of information made available through the Internet. As Simon states, "a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the over-abundance of information sources that might consume it" (1971). Further, the search and consumption of information is subject to the concept of bounded rationality, in which individuals make decisions that are not necessarily optimal as a result of a lack of available information, cognitive limitations, or time limitations (Simon, 1955). These restrictions are reflected in the concept of satisficing; just as the bird does not necessarily continue searching until it finds the world's greatest source of berries, human information foragers will stop at the information patch that provides an acceptable information sustenance to effort ratio.

Another key component of a user's foraging for information is the idea of *information scent*, defined as "the (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues such as bibliographic citations, WWW links, or icons representing the sources" (Card et al., 2001; Pirolli & Card, 1999).

Just as the foraging bird may hone in on the location of a food source based on the food's odors being carried on the wind, IT users perceive clues that help guide them to their next source of information. Within digital content environments, these clues are usually the results of design elements (choice of color, iconography) or content (word choice in links, photography) (Card et al., 2001).

Central to information foraging is the implication that, ideally, users obtain the information they need while expending the least amount of effort to obtain it. As Pirolli states, "cognitive systems engaged in information foraging … will prefer technologies that tend to maximize the value (or utility) of knowledge gained per unit cost of interaction" (Pirolli, 2007). In this sense, information foraging echoes the earlier work of Card et al., who related design effectiveness with the efficiency of completing tasks (1983). Prescriptively, then, the best possible user experience is one that optimizes time efficiency.

Within the practitioner community, this ideal has gained significant attention. Practitioner-focused usability gurus such as Jakob Nielsen and Steve Krug both argue that efficiency yields the best possible user experience, even in cases where the user is not sure what is being sought or is pursuing hedonic objectives. This, in turn, results in user behavior that they contend is optimal for the site owner, whether that be a completed purchase, increased loyalty, or an improvement in attitude toward the site's sponsor (e.g., Krug, 2009; Nielsen, 2000).

2.2.1.3. Foraging Territory and Time

In discussing and operationalizing foraging behavior as a variable, past research has conceptualized this behavior as one measured solely in terms of time spent in completing a task (Galletta et al., 2006; Pemberton, 2003; Pirolli & Card, 1999; Pirolli, 2007). In this sense, foraging refers only to the time between the beginning of the information search and the point at which the information forager is satisfied with the information acquired.

We assert, however, that foraging behavior can be measured along an additional dimension: territory. Consider again the bird flying among various shrubs in a meadow and hopping between branches in search of berries. Not only does time accrue to this activity, but during the same process the bird moves across territory in space. While it may be intuitive that the time spent in the activity and the amount of territory covered should be correlated, there may be variations depending on the bird. Some may fly faster than others. Other birds may be more adept at quickly assessing the availability of food on a given shrub. Two birds, then, may forage for the same length of time, but one may cover more territory than the other.

Similarly, two information foragers may spend the same amount of time engaged in foraging, but cover different amounts of territory. A user may view more pages and consume more information in the same amount of time if that user is more adept at making sense of information or at detecting information scent. Users may also cover more or less territory based on browsing style, for instance by having a greater or lesser propensity to click on links.

Further, the amount of territory covered may have different implications. A foraging bird who lands on more shrubs, hops on more branches, and sees more of the territory may learn more and therefore be able to eat better or more efficiently next time it visits that meadow, or may simply become more aware of the surroundings. An online information forager, similarly, may be exposed to more information and to more ideas. This exposure may likewise prove beneficial to the information forager.

In this sense, then, by adding the territory dimension to information foraging, we can reconcile the more marketing-centric idea of brand conditioning and repetition with the more information-centric concept of information foraging. In covering more territory by clicking more links, exploring more pages, viewing more images, and reading more text, a user becomes more exposed to a given website and the brand messaging that it entails.

Thus, by including this dimension of territory, we can use information foraging theory to address our research questions. To do this, we need to understand what the antecedents of information foraging are and how they affect foraging along both dimensions. Then, we can come to an understanding of how these two types of foraging affect users' feelings toward the website owner and, thus, understand what kind of behavior is optimal from the website owner's perspective.

2.2.2. Web Design Factors: Interactivity and Cognitive Absorption

The way in which a website is designed can have significant implications for the way in which a user interacts with or forages for information within that website. As such, website design be seen as an avenue for encouraging the foraging behavior most beneficial to website owners. The literature on website interactions, sourced from both information systems and marketing, proposes two principle factors of website design that can have this influence: interactivity and cognitive absorption.

Interactivity refers to the extent of information exchange on a website (Huang, 2003); it has been further described as a two-way communication-based event in which the user has some form of control and can thus modify the experience (Huang, 2003; Lowry, Romano, Jenkins, & Guthrie, 2009; Mollen & Wilson, 2010; Rafaeli, 1990; Rogers, 1986; Song & Zinkhan, 2008; Steuer, 2006). In the most basic sense, a website is inherently interactive since clicks on various site links allow a user to choose which content is communicated by the host site (Huang, 2003). Other technology applications (e.g., video games, mobile phones applications, etc.) are also clearly interactive under this definition.

In reconciling discord among a number of operationalizations and conceptualizations of interactivity, Liu and Shrum argued for the concept of interactivity consisting of three sub-dimensions (2002). These include (1) active control, the user's ability to influence communication, (2) two-way communication, which reflects bi-directional information flow, and (3) synchronicity, which refers to the speed with which a response to a request or question is given. Based on this conceptualization, a survey instrument was created and validated (Y. Liu, 2003; Song & Zinkhan, 2008) that has enjoyed use in the information systems and marketing literature (e.g., Lowry et al., 2009; Song & Zinkhan, 2008).

Research has shown that the degree of interactivity of a website can vary. The level of a website's perceived interactivity has been found to correlate with metrics of particular interest to digital content owners such as a user's intent to return to a website and user satisfaction (Koufaris, 2002; Lowry et al., 2009; Palmer, 2002). Research has therefore recommended that practitioners increase the interactivity of their websites. It is noted, however, that interactivity can be construed as a form of effort; it requires input and communication on the part of the user. As a form of effort, then, interactivity could be considered a source of cost. Indeed, tasks that are considered stressful have been shown to increase users' perceptions of interactivity (Shih, 1998). However, interactivity requires time and results in the conveyance of information, which relates to our concept of territory. Thus, we hypothesize:

> *Hypothesis 1: The greater the perception of interactivity for a site, the greater the level of information foraging both in terms of (a) territory and (b) time.*

The concept of flow and its IS-specific counterpart, cognitive absorption, have similarly been used to describe the effect of websites (as well as information technology in general) on users. *Flow* refers to a user's degree of concentrated involvement or the user's sense of "being lost" in an activity (Csikszentmihalyi, 1990); this is characterized by a seamlessness of interactions, a sense of enjoyment, and a lack of self-consciousness (Novak, Hoffman, & Yung, 2000). Flow encompasses four states of mind: flow, boredom, apathy, and anxiety (Csikszentmihalyi, 1990; Mathwick & Rigdon, 2004). The concept was introduced into the IS literature by Trevino and Webster (1992), who used it to predict user attitudes toward technology as well as level of usage. Since then, the concept has been studied extensively in the online context, particularly as they apply to online marketing. Hoffman and Novak, for instance, proposed a number of possible antecedents to flow in the online context such as skill and control, interactive speed, challenge, and focused attention, and postulated that flow would positively predict positive affect and exploratory behavior (Hoffman & Novak, 1996). Research has found that in the specific context of an online information search, flow leads indirectly to consumer attitude formation and loyalty (Mathwick & Rigdon, 2004) and that properties of flow inducement are inherent in a website (Huang, 2003).

Agarwal and Karahanna proposed a similar construct called *cognitive absorption*, which they defined as “a state of deep involvement with software” (2000, p. 673), and which they viewed as a theoretical descendent of concepts of absorption (Tellegen & Atkinson, 1974) and flow. Cognitive absorption manifests itself through five dimensions, namely temporal dissociation, focused immersion, heightened enjoyment, control, and curiosity. In addition to establishing and validating instruments to measure these dimensions and the construct itself, researchers have found that cognitive absorption associates positively with perceived usefulness and perceived ease of use of software (Agarwal & Karahanna, 2000; Lowry et al., 2013).

The idea of cognitive absorption is distinct from that of information foraging. As conceptualized under information foraging, a user has a task, intentionally follows the strongest information scent, makes rational satisficing decisions, then arrives at and gathers the desired information from the identified location. This is different from the idea of an website interaction that registers high in cognitive absorptive properties and wherein a user becomes “lost” and loses track of time. We suggest, then, that becoming detached from real-world concerns should yield a less efficient foraging experience resulting in more territory covered over a longer period of time.

Hypothesis 2: The greater the cognitive absorptive properties of a site, the greater the level of information foraging both in terms of (a) territory and (b) time.

2.2.3. Task Structure

There are a variety of intents with which users come to sites. Some of these intents are implicit in the type of site. When a user comes to the Google search engine, the typical intent is to conduct a Web search. When one comes to Amazon, shopping or product research is more likely the case. Research by Mithas et al. (2007) showed that some site design elements are more important than others in generating customer loyalty, depending on the type of site. Specifically, they found that content was more important in generating customer loyalty for information-oriented sites than for transactional sites. Conversely, they found functionality (i.e., perceived usefulness and convenience) more salient in generating customer loyalty on transactional websites. Given that these site types imply different user intents, we can construe that user intent affects the way a user perceives a website interaction.

In addition, the same site may be used by different users for different purposes (i.e., by users with different intents). Amazon.com, for instance, can be used for the sake of completing a shopping transaction, or can be used for hedonic purposes (e.g., product research for the sake of entertainment). The website for a consumer electronics manufacturer can serve users looking for pre-purchase information on a product, or can be used post-purchase by users looking for device drivers, manuals, or usage advice.

The intent of a user's visit to a website has been found to be an important predictor of the quality of the website visit. Depending on the task, its definition, and its complexity, researchers have revealed a number of important results. Jiang and Benbasat (2007), for instance, found that task complexity combined with different content types yielded different results in terms of increasing users' product knowledge. It has also been suggested that any gap between user intent and delivered experience detracts from a site's utility (Albert et al., 2004). Similarly, the fit between how a person conceptualizes a given goal and how it is presented can result in more favorable attitudes (A. Y. Lee, Keller, & Sternthal, 2010). Novak et al. (Novak, Hoffman, & Duhachek, 2003; Novak et al., 2000) studied the impact of intent on the construct of flow within the online context. They found that the level of goal-focus in the user's intent significantly predicted the amount of flow experienced by users.

For the purpose of our data collection and in accordance with a significant body of past literature on information search (e.g., Browne et al., 2007; Davies, 2003; Pirolli, 2007; Simon, 1973), we conceptualize the user's intent in terms of task structure. Task structure can be defined as "the degree to which the necessary inputs, operations on those inputs, and outputs are known and recognizable to the decision maker" (Browne et al., 2007, p. 92). In our study, we differentiate high task structure from low task structure. A task with high structure is likely to have a single correct answer that does not require significant abstraction of thought; we note that, in this sense, a highly structured task is likely to be low in terms of complexity and likely correspond with Novak et al.'s concept of goal-focused intent (2003, 2000). A task with low structure, on the other hand, is likely to be more ambiguous and may therefore have several reasonable answers of varying quality among which the individual may have to make evaluations. A low-structure task, then, should be more similar to Novak et al.'s concept of experiential intent (2000; 2003).

We expect that in a more closed-ended task (i.e., a task with higher task structure), site visitors will have a more focused approach to their information search. Information scent should be more salient and easier to detect when the task to be completed involves less complexity; users should have a better idea what they seek and thus find it easier to detect the appropriate scent. Conversely, visitors performing a more open-ended task (i.e., a task with lower task structure) will need more time to pick up the appropriate information scent and perform sense-making activities. Thus:

> *Hypothesis 3: The lower the task structure, the greater the level of information foraging both in terms of (a) territory and (b) time.*

2.2.4. The Dependent Variable: Brand Engagement

Finally, given organizations' expenditures, both in terms of cost and effort, in creating websites, it is crucial that the purpose of those sites be understood. As mentioned, within the context of this study, we have identified the objectives of brand-focused websites as our focus. We therefore identify the concept of brand engagement as our dependent variable. Here, we define *brand engagement* as the combination of user awareness of and involvement with a brand and the commitment to act based on that awareness and involvement (Keller, 2008). The three components of our conceptualization echo those proposed by Lavidge and Steiner in their discussion of the purpose of advertising, namely the cognitive (knowledge and awareness), emotional (attitude and affect), and motivational (willingness to act) (1961).

In their review of the general concept of engagement, Mollen and Wilson (2010) sought to reconcile the concept as understood by practitioners against the academic views on flow and interactivity. In doing so, the authors describe engagement as both a cognitive and affective commitment to a relationship with a brand or product. The concept of brand engagement that we use here hews more closely to that understood by practitioners and, in so doing, builds from the marketing-based conceptualization of engagement (Keller, 2008). We thereby differentiate our concept from those of others who consider engagement as primarily a psychological state (e.g., Brodie, Hollebeek, Juric, & Ilic, 2011).

Similar to related concepts, we propose brand engagement as residing on a single continuum. Interactions with a given stimulus (such as a brand-sponsored website) can thus result in an increase or decrease in the user's level of brand engagement. By level of commitment we refer to the likelihood of a user taking action(s) favorable to the firm hosting the site or owning the brand; for example, the likelihood of purchasing a product or the likelihood of advocating on behalf of the brand. This involvement thus reflects both cognitive and emotional investment.

Given this conceptualization of brand engagement, we now discuss operationalizing the construct. An underlying theoretical assumption is that the presence of engagement can best be identified through inference based on a change in users' attitudes and expected behaviors as the result of interaction with a stimulus.

Because most online advertising stimuli lead users to a website and since websites are the core focus for firms' online branding efforts, we consider a firm's website to be a reasonable setting in which to study brand engagement. Accordingly, we infer a change in engagement based on three readily operationalized measurements: (1) the change in a user's attitude toward a company's brand; (2) the change in a user's intent to purchase; and (3) the change in the user's product knowledge. Again, this threefold approach echoes that found in the marketing-focused advertising literature (e.g., Lavidge & Steiner, 1961).

In predicting how user foraging behavior affects changes in a user's sense of brand engagement, we return to our initial ideas of brand conditioning and information foraging. As discussed, under brand conditioning, the more a person is exposed to an idea, the stronger the relationship with that idea becomes. Thus, we expect that the more territory covered on a website, the greater the increase in the user's brand engagement.

> *Hypothesis 4: The greater the user foraging behavior in terms of territory, the greater the increase in user brand engagement in terms of brand attitude, purchase intent, and product knowledge.*

On the other hand, considering the work in information foraging showing the positive relationship between efficient information gathering experience measured in terms of time and positive feelings toward the sponsoring site, we hypothesize that time spent foraging will have the opposite relationship. Thus:

> *Hypothesis 5: The greater the user foraging behavior in terms of time, the lesser the increase in user brand engagement in terms of brand attitude, purchase intent, and product knowledge.*

2.2.5. Consolidated Hypothesis Model

Assembled together, the hypotheses above yield the model shown in Figure 2.1. The next section details the methodology that was used to gather data for testing these hypotheses.

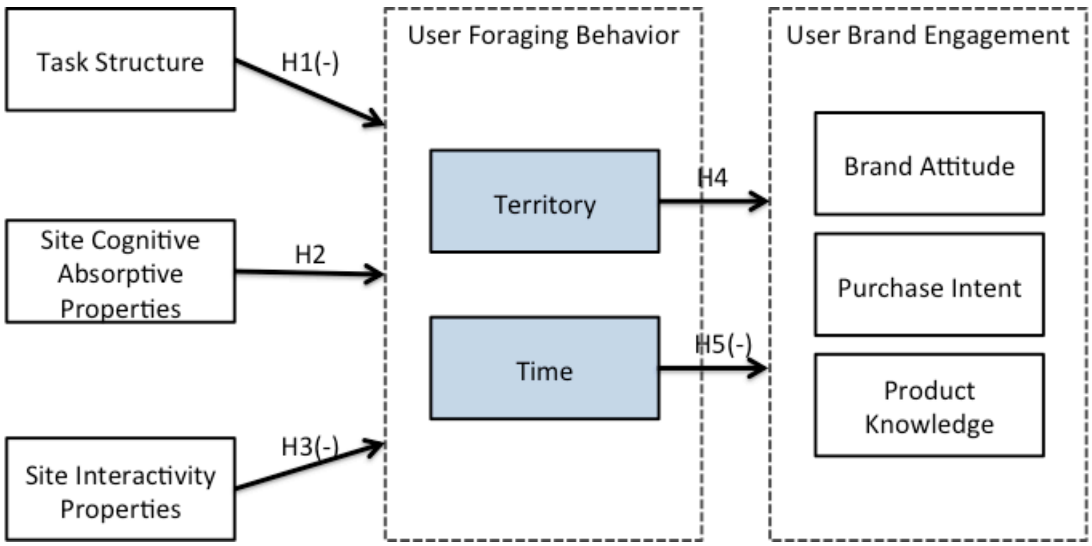


Figure 2.1. Consolidated hypothesis model.

2.3. Data Collection

Data were collected in a laboratory setting from a total of 254 participants recruited from an undergraduate psychology participant pool at a major university in the Mid-Atlantic region of the United States. Participants received academic credit for participating, but no other compensation was awarded. Of these 254 participants, 36 were part of a control group given a dummy task to evaluate any Hawthorne Effect issues that may have come about as a result of the laboratory setting (none were detected). In addition, 29 subjects submitted incomplete data (e.g., failed to complete the task assigned) and were omitted from the data set, leaving a relevant sample of 189 observations.

Participants were asked to interact with website stimuli so that we could track and measure their activities and then later survey participants to understand the impact of those activities (please see Appendix A for explication of survey instruments and descriptive statistics). In determining the appropriate stimulus website, it was important to select a product category well known to our participant pool to ensure that they would have a natural interest in the product type. Additionally, a high-involvement product type increases the likelihood of identifying significant relationships since such products have been found to induce higher levels of "play", which has been found to increase positive attitudes and loyalty (Mathwick & Rigdon, 2004). Such a product would thereby make it more probable to detect cognitive absorption among participants. Thus, we selected car manufacturer websites as our stimuli as cars are likely to be well known to our participants and are a high-involvement product (Johnson & Russo, 1981). In this, we follow the example set by other highly cited papers, which have used this product category as a context for study based on similar rationales (e.g., Moorthy, Ratchford, & Talukdar, 1997; Punj & Staelin, 1983; Srinivasan, Narasimhan & Ratchford, Brian T., 1991).

To test our hypotheses regarding site design in terms of cognitive absorptive properties and interactivity properties, we needed a range of values for these two variables and, thus, we included stimulus sites from three different car manufacturers. Furthermore, to counter the potential bias that might be introduced due to difference in brand associations, we controlled for these by selecting websites from brands that do not necessarily have strong, specific brand positions. We therefore identified Toyota's US website, Mitsubishi Motors' US website, and Seat Motors' UK website (Seat, a Spanish car manufacturer owned by Volkswagen, does not sell cars in the US) as appropriate website stimuli[2]. To account for the relative familiarity of these brands and the potential impact that might have on results (Baker et al., 1986; M. C. Campbell & Keller, 2003; R. J. Kent & Allen, 1994; Laroche, Kim, & Zhou, 1996), we controlled for brand familiarity in our estimations. Further, to minimize bias against Seat during the task, participants were asked to assume that Seat planned on selling cars in the US market within the next six months.

Prior to beginning the task, participants were required to complete a pre-task survey. This survey included general items regarding demographics (length of US residency, age, gender, student status, and number of semesters studied) and car driver/ownership status (whether the participant had a driver's license, owned a car, or planned to buy or lease a car in the foreseeable future). Length of US residency was specifically included to control for differences in brand impression across different parts of the world, where the companies' advertising campaigns and other factors may have resulted in country-based differences in brand perception (we note that 97.3% of

[2] This determination was based in part on website popularity values as found at Alexa.com on October 4, 2013. According to Alexa, Toyota's USA website was the most frequently visited Japanese auto maker website in the US, while Mitsubishi's was the least visited. Similar findings emerged from a pilot study (n = 40), where, among brands polled, Toyota was in a virtual tie with Honda as the most familiar Japanese car-maker brand (5.89 mean value on a seven-point Likert scale) and Mitsubishi the least familiar (2.90). In the same study, Seat received a mean familiarity score of 1.32. Familiarity in this pilot study was based on a five-item instrument consisting of the first five items listed for brand familiarity in Appendix A, Table A.1.

participants responded that they had lived in the US for 10 years or more). In addition, we measured participants' *product involvement*, or the degree to which an individual engages with a product, using Zaichkowsky's (1994) established product involvement instrument. We took this measurement for use as a control variable.

In addition, participants completed instruments to measure Brand Purchase Intent (Coyle & Thorson, 2001; Kim & Biocca, 1997), Brand Attitude (Gardner, 1985; Yuping Liu & Shrum, 2002; Machleit & Wilson, 1988), Product Knowledge, and Brand Familiarity (Alba & Hutchinson, 1987; M. C. Campbell & Keller, 2003; Collins, 2007; R. J. Kent & Allen, 1994; Lowry et al., 2008; Stewart, 1992). Brand Purchase Intent and Brand Attitude are established instruments. Product Knowledge was measured by asking participants to name up to five car models manufactured by each of the three stimulus brands and, in this sense, is an objective measure. As used here, Brand Familiarity represents a new instrument and was constructed through seven items derived from existing literature. This construct was found to have sufficiently high convergent validity, Cronbach's α=0.898 (Nunnally & Bernstein, 1994). We also note that, with the exception of brand familiarity, participants were asked to complete each of these instruments three times, once prior to the task and once immediately after the task, enabling us to create difference scores for the brand engagement measures. Actual items used for these constructs can be found in Appendix A, Table A.1.

Prior to beginning the task, participants were randomly assigned to either the high structure task (n=98) or the low structure task (n=91) condition. In the high structure task condition, participants were instructed to use the three stimulus websites (and only those websites) to answer each of nine focused, factual questions, three questions for each brand. Questions were clearly labeled by brand, e.g., each of the three questions that required use of the Toyota site were labeled as being about Toyota and included the URL for the Toyota website. Each question was multiple-choice, with ten options per question and an additional “Can’t Find/Don’t Know” option. See Appendix A, Table A.3 for questions used in the high structure condition.

Participants in the low structure condition were given the following instruction:

> *Using the websites at Toyota.com, mitsubishicars.com, and seat.co.uk, find the four models that you find the most appealing (i.e., that you would most likely want to buy if you were to buy a car within the next five years) and list to reasons why each model is appealing to you. Please limit your selections to vehicles you find on these three websites. You have up to five minutes to complete this task.*

Foraging behavior was measured in terms of time (in seconds) and territory (number of pages viewed). These observations were calculated through the use of the Windows Problem Steps Recorder (PSR) software, which captures all user inputs in an XML-formatted log file (see Appendix A, Figure A.1). Each participant thus generated a unique file that could be tied back to the matching survey answers. These XML files were then parsed using a custom-made Perl script to yield a tab-delimited file (see Appendix A, Figure A.2), which was then parsed further to determine values for time and territory for each participant for each of the three sites included as part of the task.

Participants were further randomly assigned into one of three groups, Toyota, Mitsubishi, or Seat. This assignment became important following completion of the task, when participants were asked to complete additional survey items applicable exclusively to the one brand to which they were assigned. As mentioned, the constructs measured included Brand Purchase Intent, Product Knowledge, and Brand Attitude. In addition, participants were asked to complete instruments for both Cognitive Absorption (Agarwal & Karahanna, 2000) and Interactivity (Y. Liu, 2003; Song & Zinkhan, 2008) for the site of the brand to which they were assigned. While all participants completed their assigned task using all three sites, each participant answered post-task survey questions exclusively for the one brand to which they were assigned. As such, there was no difference in the task completed based on assigned brand; the only difference was found in the post-task survey.

The cognitive absorption instrument used was nearly identical to that in Agarwal and Karahanna (2000), in which the construct is considered to exist in five sub-dimensions of temporal disassociation, immersion, enjoyment, control, and curiosity. To minimize the possibility of survey fatigue, however, we used only the three highest loading items (as reported by Agarwal and Karahanna) for each sub-dimension. All sub-dimensions had sufficient reliability based on Cronbach's α (Nunnally & Bernstein, 1994); temporal disassociation sub-dimension $\alpha = 0.847$, immersion sub-dimension $\alpha = 0.757$, enjoyment sub-dimension $\alpha = 0.868$, control sub-dimension $\alpha = 0.724$, curiosity sub-dimension $\alpha = 0.852$.

The interactivity instrument used was that created by Liu (2003) and later adapted by Song and Zinkhan (2008). This instrument conceptualizes Interactivity into its three sub-dimensions of control, two-way communication, and synchronicity. Again in the interest of mitigating possible bias caused by survey fatigue, we abbreviated this instrument by selecting the three items for each sub-dimension which showed the highest loadings in past use (Song & Zinkhan, 2008). Again, all sub-dimensions had sufficient reliability based on Cronbach's α (Nunnally & Bernstein, 1994): control sub-dimension $\alpha = 0.784$, two-way communication sub-dimension $\alpha = 0.803$, and synchronicity sub-dimension $\alpha = 0.885$. Here, factor loadings are acceptable, with measures of each of the three sub-dimensions loading most strongly with the correct other measures. See Appendix A, Table A.8 for items used to form the Cognitive Absorption and Interactivity constructs.

2.4. Analysis

Mean values by task type and brand (implying familiarity) are shown in Table 1 below. Note that the columns for product knowledge, purchase intent, and brand attitude reflect delta scores, the mean increase (or, in the case of negative values, decrease) for each variable as a result of the participants' completion of the experimental task.

Task	***n***	**Pages**	**Time**	**Time/Pg**	**Know Δ**	**Purch Δ**	**Att Δ**
Structured	98	19.95	147.41	7.39 s	0.735	0.126	0.359
Mitsubishi	35	23.69	171.32	7.23 s	0.914	-0.038	0.343
Seat	32	24.97	168.16	6.73 s	0.438	0.583	0.578
Toyota	31	15.81	99.00	6.26 s	0.839	-0.161	0.151
Unstructured	91	17.25	119.63	6.94 s	1.209	0.271	0.672
Mitsubishi	32	17.91	129.61	7.24 s	1.125	0.322	0.563
Seat	31	21.48	129.81	6.04 s	1.129	0.581	0.887
Toyota	28	13.32	96.96	7.28 s	1.393	-0.131	0.559
Total	**189**	**18.65**	**134.03**	**7.19 s**	**0.962**	**0.196**	**0.510**

Table 2.1. Mean values by task type and treatment group. Time measured in seconds. Know = Product Knowledge (scale of 0-4), Purch = Brand Purchase Intent (1-7), Att = Brand Attitude (1-7).

Data were then analyzed in two steps. In the first step, a regression examined the effects of the antecedents on user foraging behavior (i.e., Hypotheses 1-3), relating task structure and website design properties to both foraging time and foraging territory. In the second step, a regression evaluated the relationships between foraging time and territory and brand engagement (Hypotheses 4 and 5). Using the Baron & Kenny method (1986) and the Sobel test statistic, mediation was insignificant at the 0.05 level.

Given the potential for correlation of error terms in each of these steps, seemingly unrelated regression estimation (SUR) was used to analyze each step as a system of equations (Zellner, 1962). The first step was specified using Equation (1) for Foraging Territory and Equation (2) for Foraging Time. Note that, under SUR, the two equations were evaluated simultaneously.

Equation 1:
Foraging Territory = $\beta_0 + \beta_1$ *(Task Structure)* + β_2 *(Cognitive Absorptive)* + β_3 *(Interactivity)* + β_4 *(Product Involvement)* + β_5 *(Brand Familiarity)* + β_6 *(Age)* + β_7 *(Gender)* + β_8 *(Residency Status)* + ε

Equation 2:
Foraging Time = $\beta_0 + \beta_1$ *(Task Structure)* + β_2 *(Cognitive Absorptive)* + β_3 *(Interactivity)* + β_4 *(Product Involvement)* + β_5 *(Brand Familiarity)* + β_6 *(Age)* + β_7 *(Gender)* + β_8 *(Residency Status)* + ε

The results of this estimation are shown in Table 2.2 (correlations among all included variables are shown in Table A.9 in Appendix A).

	Territory	**Time**	
Task Structure	-0.62***	-0.65***	H1(-)
Cognitive Absorptive	0.12**	0.12**	H2(+)
Interactivity Properties	-0.16**	-0.15**	H3(-)
Brand Familiarity	*0.09***	*0.14***	
Product Involvement	*-0.04*	*-0.03*	
Age	*0.01*	*-0.01*	
Gender (Female)	*0.11***	*0.05*	
US Residency	*0.15***	*0.12***	
n	189	189	
Adjusted R^2	0.37	0.41	
Model χ^2	113.41***	132.28***	

Table 2.2. Results of first regression. *** = p < 0.001, ** = p < 0.01, * = p < 0.05; standardized coefficients reported.

In this first-step regression, we find support for all three hypotheses tested. Greater task structure results in less foraging both in territory and time (H1); the greater the perceived cognitive absorptive properties of a site, the greater the foraging behavior in terms of both territory and time (H2); and the greater the interactivity properties of the site, the lower the user foraging behavior in terms of territory and time (H3). Estimations significantly predict both foraging behavior variables[3].

In the second step, hypothesized relationships between user foraging behavior and changes in brand engagement were tested using the following system of equations.

Equation 3:
Δ *Brand Attitude* = $\beta_0 + \beta_1$ *(Foraging Territory)* + β_2 *(Foraging Time)* + β_3 *(Task Structure)* + β_4 *(Cognitive Absorptive)* + β_5 *(Interactivity)* + β_6 *(Product Involvement)* + β_7 *(Brand Familiarity)* + β_8 *(Age)* + β_9 *(Gender)* + β_{10} *(Residency Status)* + ε

Equation 4:
Δ *Purchase Intention* = $\beta_0 + \beta_1$ *(Foraging Territory)* + β_2 *(Foraging Time)* + β_3 *(Task Structure)* + β_4 *(Cognitive Absorptive)* + β_5 *(Interactivity)* + β_6 *(Product Involvement)* + β_7 *(Brand Familiarity)* + β_8 *(Age)* + β_9 *(Gender)* + β_{10} *(Residency Status)* + ε

Equation 5:
Δ *Product Knowledge* = $\beta_0 + \beta_1$ *(Foraging Territory)* + β_2 *(Foraging Time)* + β_3 *(Task Structure)* + β_4 *(Cognitive Absorptive)* + β_5 *(Interactivity)* + β_6 *(Product Involvement)* + β_7 *(Brand Familiarity)* + β_8 *(Age)* + β_9 *(Gender)* + β_{10} *(Residency Status)* + ε

[3] We also note here a significant, positive relationship between the control variable of brand familiarity and our two foraging variables. We are aware of no current literature that theoretically or empirically relates brand familiarity with foraging behavior. Given voluminous literature relating it to elements of brand engagement (e.g., brand attitude, purchase intent, and advertisement memorability) (e.g., Baker, Hutchinson, & Moore, 1986; M. C. Campbell & Keller, 2003; Johnson & Russo, 1981; R. J. Kent & Allen, 1994; Laroche, Kim, & Zhou, 1996; Machleit & Wilson, 1988), it was deemed important to include this as a control variable in the second regression (predicting brand engagement). Since it was included in the second regression, the variable was likewise included in the first. However, based on the significant finding here, future researchers may want to consider the role of brand familiarity in predicting user foraging behavior. It may be that, even in cases of a high structure task, a user's familiarity with a brand increases that user's interest in the brand and, thus, results in users opting to conduct a greater amount of foraging information related to its products. In these cases, it is possible that user intent becomes less structured as a result of inherent interest.

Equations (3), (4), and (5) were likewise estimated simultaneously using SUR, and the results of this second step are given in Table 2.3 below (correlations among included variables are shown in Table A.10 in Appendix A).

	Δ Brand Attitude	Δ Purchase Intent	Δ Product Knowledge	
Foraging Territory	**0.23****	**0.26****	**0.23****	**H4(+)**
Foraging Time	**-0.22****	**-0.27****	**-0.22****	**H5(-)**
Task Structure	***-0.12*****	***-0.10***	***-0.11****	
Cognitive Absorptive	***-0.06***	***0.04***	***-0.06***	
Interactivity	***-0.06***	***-0.09***	***-0.06***	
Brand Familiarity	***0.17*****	***0.12****	***0.16*****	
Product Involvement	***-0.01***	***0.06***	***-0.02***	
Age	***0.26*****	***0.12***	***0.25*****	
Gender (Female)	***-0.01***	***-0.05***	***-0.01***	
US Residency	***0.02***	***-0.07***	***0.02***	
n	**189**	**189**	**189**	
Adjusted R^2	**0.24**	**0.24**	**0.23**	
Model χ^2	**34.71*****	**30.00*****	**38.93*****	

Table 2.3. Results of second regression. *** = p < 0.001, ** = p < 0.01, * = p < 0.05; standardized coefficients reported.

Again in this second step, our analysis supports both of our hypotheses. The greater the foraging time, the lesser the increase in brand engagement in terms of brand attitude, purchase intent, and product knowledge (H4). Also, the greater the foraging territory, the greater the increase in brand engagement in terms of brand attitude, purchase intent, and product knowledge (H5). We again find that the estimated models significantly predict our dependent variables. Again, using the Baron & Kenny method (1986) and the Sobel test statistic, mediation was insignificant at the 0.05 level.

2.5. Discussion

2.5.1. Findings and Implications

Both steps of our analysis yield interesting results. We find that users coming to a website with a higher task structure in mind tend to forage less both in terms of time and territory. We also find that sites that are perceived to have higher cognitive absorptive properties result in users who forage more (both in time and territory), while those that are perceived to be more interactive result in users who forage less. Therefore, if both kinds of foraging were either good or bad, we would have a clear recommendation for designers and website owners.

However, as hypothesized, this is not the case. Instead, we find that, while foraging behavior significantly influences brand engagement (in terms of brand attitude, purchase intent, and product knowledge), it does so in opposite ways depending on the dimension of foraging behavior. An increase in foraging time, as expected from the information foraging literature, results in a smaller improvement in brand engagement. An increase in foraging territory, however, as expected from brand conditioning and repetition, results in a larger improvement in brand engagement. Thus, the best result could be seen where users are exposed to more of a website's territory, but complete their exposure in a shorter period of time.

We present this as a significant challenge to designers and site owners. How can users be influenced through design so as to encourage them to forage a larger territory in less time? There are clear implications for potential technical issues that may cause latency, which increases foraging time without any increase in foraging territory. However, design may also be key in addressing this need by better executing the concept of information scent. Further, different interactivity mechanisms, such as new navigation techniques, innovative use of voice or video, etc. may have a positive impact. Finally, there may be meaningful ways to decompose both interactivity and cognitive absorption to identify specific sub-dimensions that lead to optimal foraging. Each of these possibilities requires further research, which can be based on the findings presented here.

While this study was conducted within the context of brand-focused websites, we suggest that our findings should be generalizable to other websites where brand development is important. We further suggest that brand engagement is important for almost all sites. While some web properties may exist primarily for the sake of e-commerce (e.g., Amazon), search (Google), social networking (Facebook), or other use cases, developing positive brand associations with customers remains an important factor in determining the success of these site types as well.

2.5.2. Limitations and Future Research

While this research presents important findings, we note there are also some limitations to our approach. First, our study uses a simple representation of territory and does not, for instance, explore factors such as website depth or navigation patterns. We expect, however, that a more complex representation would add to the significance of our findings and is worthy of future research.

Second, it has been argued that Cognitive Absorption is a construct that may be better evaluated through its sub-dimensions as, according to Flow theory (Csikszentmihalyi, 1990) upon which it is based, the sub-dimensions do not necessarily occur all at one time (Lowry et al., 2013). In their study on hedonically motivated adoption, Lowry et al. operationalize Cognitive Absorption as four constructs, with curiosity, joy (enjoyment), and control antecedent to immersion, which the researchers consider a merged form of temporal dissociation and focused immersion (2013). Given this alternative conceptualization, it may be useful for future research to test a similar model using this less monolithic view of Cognitive Absorption. This approach could help further explain the curious hypothesis and result in which Cognitive Absorptive properties and Interactivity properties of a website have opposite effects on user foraging behavior despite the two constructs sharing a common sub-dimension (Control), which is measured using similar instrument items (see Table A.8 in Appendix A for a listing of instrument items).

Third, the study looks at the question of optimal website user behavior only in the context of branding-specific websites and only within the automobile product category. While we feel that our findings should be generalizable to branding objectives on other types of sites or in other product categories, this cannot be definitively established without replicating this study in those contexts.

Finally, this study focuses on a "traditional" website encounter using a PC and browser. Again, as information foraging behavior can be found on other platforms (e.g., smartphones, video game consoles) and as content sites built specifically for use on those platforms nevertheless have branding objectives, we expect that our findings will generalize to these other devices and means of accessing content. Still, other device-specific design mechanisms (e.g., touch interfaces) cannot be accounted for in this study and therefore merit further research.

2.5.3. Conclusion

This study constitutes a meaningful step forward in understanding website foraging behavior and, in particular, how it relates to optimal outcomes for website owners. It makes three major contributions. First, we bring the idea of brand engagement as an information systems-relevant variable into the IS literature. It joins other, more established variables such as intent to purchase and intent to re-visit as an important dependent variable for further study. Second, we reconcile the traditional marketing perspective of brand conditioning and repetition with that of information foraging by introducing the idea of a second foraging dimension, territory. Our experimental results indicate that website designers need to consider both time and territory in order to optimally influence a user's brand engagement during a website visit. Finally, we present a significant challenge for website designers and owners by showing that optimal user behavior requires the covering of both more territory as well as more time-efficient task completion. This challenge calls for further research on design mechanisms that enable website users to forage more information in a shorter period of time.

Predicting the Intent of Sponsored Search Users: A User Session-Level Analysis

3.1. Introduction

Website-owning firms face a difficult problem. These firms invest considerable time and money into building and maintaining websites (see Chapter 1). Meanwhile, past research has shown that the success of those websites, as determined by a number of outcome constructs, can depend on the intent of users arriving on the site. These outcome constructs include loyalty (Mithas et al., 2007), satisfaction (Kohli, Devaraj, & Mahmood, 2004; McKinney, Yoon, & Zahedi, Fatemeh Miriam, 2002), trust (M. K. O. Lee & Turban, 2001), flow experience (Novak et al., 2003), attitude (A. Y. Lee et al., 2010), product knowledge (Zhenhui Jiang & Benbasat, 2007), and foraging behavior (Browne et al., 2007; Dunn, Ramasubbu, Galletta, & Lowry, 2014). Knowing a user's intent, however, is a difficult proposition necessitating some degree of programmed, algorithmic clairvoyance.

Indeed, while knowing what the user has in mind for a given visit may have substantial importance, in most cases website owners have only limited information — referring site and IP address — from which to infer this intent. Is it possible for website owners, then, to programmatically infer meaningful intent cues based on information associated with the route that users took to arrive there? And, if so, how?

Investigating this possibility requires a significant amount of data. Further, these data should allow us to control for the almost infinite variety of ways that a user can enter a website. Therefore, a data set from a single website traffic source makes a strong candidate for study. To this end, we have identified sponsored search traffic data as a useful data source to explore the potential for making meaningful inferences regarding website visitor intent.

Commercial website owners spend considerable money not only to build and maintain their sites, but also to execute tactics intended to attract users to interact and transact with those sites. One of the most popular tactics for building traffic is sponsored search keyword advertising, in which advertisers pay a per-click fee to attract users based on specified search queries. It is also the online traffic attainment strategy on which companies spend the most money, with an estimated $53.6 billion to be spent worldwide in 2014 (Lunden, 2013) and $21.7 billion to be spent in the US alone (eMarketer, 2012). These expenditures further demonstrate the value in knowing more about the user at the time of website entry and a correspondingly high potential cost of ignorance. Interestingly, these users arrive with an additional piece of potentially useful information: the search term that generated the link to the website. Further, users clicking through on such links can be tracked through "cookies", and, over time, their searches can form a user search session such that the same user's searches over time can be monitored and evaluated. We assert that these search sessions provide a potentially fertile testbed for discovering the ability to infer user intent based on information related to the user's means of arrival.

To this end, we have acquired a very large dataset of sponsored search click data from a major travel agency in Asia. The data include user identifiers (IP addresses), the search terms used that resulted in clicks, time stamps for these clicks, and one key user outcome for the resulting visit (whether or not the user purchased). This data set includes information regarding over 2 million query-click dyads over a 12-month period. The size of the dataset not only provides substantial territory for exploration, but also is indicative of the quantity of data with which advertising firms must contend in order to optimize their advertising campaigns.

To understand how these data can be used to predict user intent, we first explore the range of user intent possibilities. The marketing literature has given us a stream of research focused on the purchase funnel, the process by which a consumer, whether online or offline, initially considers a range of possible product or service offerings, then, by acquiring and processing information, narrows down the possible alternatives before finally purchasing the best suited choice (e.g., Kotler, 1997; Lavidge & Steiner, 1961; E. S. E. Lewis, 1903). Given the commercial nature of our dataset, this marketing approach seems particularly relevant. We note, however, that this funnel approach coincides with formulations of the broader problem-solving literature (e.g., D. J. Campbell, 1988; Simon, 1960, 1981). These perspectives enable us to identify user intent based on location in the funnel or in terms of structure type.

Prior to analyzing our data, however, we must first also create a way to categorize the data included in the query-click dyads. For this, we develop a novel typology that enables us to view search query clicks in terms of the breadth and specificity of the keyword terms used. Under this typology, both breadth and specificity are determined to have multiple sub-dimensions. The resulting typology is then applied to the query-click dyads.

Further, we acknowledge that a user’s intent may be better expressed over the course of multiple queries and clicks. With this in mind, we introduce a new unit of analysis to the sponsored search literature, the user search session, which refers to all the query-click dyads that can be ascribed to a single user intent (e.g., to a user’s intent to book a trip to New Zealand).

With the range of intent in place and user search session identified as our unit of analysis, we can then explore this massive data set. We do this by first establishing basic hypotheses regarding the implications of the depth (specificity) and breadth (number of categories referenced) of search sessions. These are then tested to determine their ability to predict user intent (i.e., location within the purchase funnel). We then build on these analyses by exploring the potential for categorizing search sessions as a potential means for algorithmically determining user intent upon arrival. Following the analyses, we discuss our findings and their implications both for research and practice.

3.2. Background

3.2.1. The Sponsored Search Context

3.2.1.1. Sponsored Search Overview

Due to its centrality to this research, we offer here a brief overview of sponsored search advertising, particularly as it applies to the website-owning firm (i.e., the advertiser)[4]. Sponsored search advertising has been in existence since its introduction in 1998 by GoTo.com and has become the primary source of Google's revenues since being adopted as part of that company's search engine in 2002 (Jansen & Mullen, 2008). In its most basic sense, sponsored search advertising works similarly to other forms of advertising such as television or radio; search engines attract an audience (users) by providing content, then sell access to that audience to advertisers (website-owning firms). In the search context, this access occurs when a user performs a search query and the engine returns a search results page that includes both paid advertisements as well as "natural", algorithmically derived search results. In general, the paid advertisements are separated and designated in some way; Google, for instance, places paid advertisements at the top and right side of the search results page with the word "Ad" included either as part of the paid listing or at the top of a series of paid listings (see Figure B.1. in Appendix B).

However, there are a number of facets of sponsored search advertising that make it very unlike traditional advertising forms. First, the advertising is not solely based on

[4] For a more in-depth review of sponsored search advertising, see Jansen and Mullen (2008).

decisions made by the advertiser and the advertising platform. Instead, the user is a direct participant in determining which advertisements are shown. In sponsored search, advertisers select the terms for which they would like their advertisements to be relevant as well as the content of the advertisement they would like to have shown on relevant queries' search results pages. Then, when a user performs a search for a term targeted by an advertiser, that advertiser's ad (as well as others similarly targeted) is displayed. For instance, a company selling home theater projectors may choose to advertise on the query "high definition projector". When a user of the search engine then performs a search for "high definition projector", the projector company's ad is shown. Because advertisers have some notion regarding the user's subject area of interest, as divulged through the search query, there is the potential for a strong match between the ad's content and the user need that prompted entry of the query.

These ads consist of both a link and advertising copy. Most commonly, this link is set by the advertising firm to direct the user to a page hosted on the advertiser's website, while the copy attempts to entice (qualified) users to click. This brings us to another important and unique feature of sponsored search advertising: advertisers only pay the search engine when their ads receive clicks. In other words, unlike the television market, where the network charges for the mere presentation of an ad, in sponsored search advertising, payment only occurs when the user takes action to access the advertiser's website. For the hypothetical home theater projector company mentioned above, if a user searches for "high definition projector", sees the company's advertisement and reads its message, but opts to click on an unrelated link, the home theater projector company pays nothing.

The advertiser is required to pay the search engine a fee, however, when the user opts to click on the home theater projector company's sponsored search link. The amount of this fee is determined by the final unique facet of sponsored search advertising discussed here: its pricing mechanism. For competitive search query terms (i.e., those for which more than one advertiser opts to present an advertisement), search engines generally display multiple ads. These several ads are displayed on a search results page in a rank order based on a special form of generalized second-price auction. Under a generalized second-price auction, bidders submit secret bids; the highest bidder wins the auction, but pays only a small premium (e.g., one cent) more than the second-highest bid (Edelman, Ostrovsky, & Schwarz, 2007). This process is used to determine the order of advertisements displayed on a search results page. Advertisers establish a per-click bid for a given search term (e.g., "high definition projector"). The highest bidder gets the first position, the second-highest the second, etc. If an advertiser's link is clicked, that advertiser then pays one cent more than the next-highest bid (Jansen & Mullen, 2008).

In determining higher and lower bids, we note that modern search engines consider more than just the monetary bid submitted by each advertiser. Google, for instance, also includes at least two additional factors: (1) an algorithmically-derived "quality score" for each advertiser based on its site's content and (2) the specific ad's history of attracting clicks, whereby ads that have historically received more clicks (i.e., have a higher click propensity) are assigned a higher bid value. With this complex approach to determining bids for comparison, when an advertiser's ad is clicked, that advertiser is charged an amount equivalent to the quality- and click propensity-adjusted bid they would have needed to be found equivalent to the next lower bidder, plus one cent.

3.2.1.2. Sponsored Search Terminology and Metrics

The unique sponsored search approach to advertising results in a multitude of metrics and terminology used by advertisers to manage sponsored search advertising campaigns. These have been used in past research and are useful in addressing our research questions as well.

Within a search engine (e.g., Google), a user will conduct a *search* by entering one or more *keywords*, or terms that the user chooses to describe the content sought. The set of keywords taken together (e.g., "high definition projector") form a *search query*. Once the user has entered the query, the search engine accesses its data and algorithmically determines the sites most likely to deliver the sought-after content. These links are presented to the user on a search results page.

Looking from the advertiser's point of view adds additional terms and metrics. The term *impressions* refers to the number of times that a given advertisement was displayed to a user. The number of times that a user clicked on an ad is referred to as *clicks*. Advertisers often evaluate the performance of ad and advertisement-query dyads through the ratio of clicks to impressions, also known as the *click-through rate*. In addition, performance may also be evaluated based on the average cost paid for a click, called *cost-per-click* (CPC).

Finally, organizations will measure success against the objective of their campaign. An industry report has shown that the three most commonly cited objectives for sponsored search advertising are direct sales, lead generation, and branding (Econsultancy.com, 2011). Organizations often manage their sponsored search campaigns by tracking ads' effectiveness at achieving these objectives. Sales can be measured through actual sales transacted through the website as a result of a click and, similarly, leads can be measured through actual contact information submitted to the company (a user submitting a loan application for instance, or a corporate customer submitting an information request to a potential vendor). Both a completed sale and a generated lead can be considered forms of *customer conversion*; an opportunity to interact with a user is converted into a measured action. These conversions can be used to form another metric, *conversion rate*, or the ratio of conversions to clicks[5]. Companies seeking branding or brand engagement objectives (Dunn et al., 2014; Keller, 2008; Mollen & Wilson, 2010) will often measure success based on activities the user conducted after clicking on an advertisement (e.g., number of pages viewed or time spent on site). Where website owners better understand the intent of users arriving on their sites via sponsored search, we expect they will be better able to optimize these campaigns according to these metrics.

---

[5] We note also the use of revenue- and profit-to-cost ratio metrics in many organizations conducting sponsored search campaigns, for instance "revenue per dollar", which compares revenue generated by clicks to the cost of those clicks.

3.2.2. Search-Related Research

3.2.2.1. Overview

Much of the research conducted on sponsored search has come from the technical or modeling perspective. These studies have often looked at economic effects underpinning the decisions of and outcomes for sponsored search participants. Such studies have researched topics including the relationship between an advertisement's rank and click-through rate and conversion rate (Ghose & Yang, 2009), auction inefficiencies (Edelman & Ostrovsky, 2007), potential auction equilibria (Edelman et al., 2007), auction simplification mechanisms (Milgrom, 2010), possible welfare distribution effects of alternative search engine policies (Yao & Mela, 2011), and effects of a regulated sponsored search market on possible adverse selection (Animesh, Ramachandran, & Viswanathan, 2010). While this stream of research has provided useful insight into the underlying mechanisms of sponsored search and suggests important implications in particular for the search engines themselves, it does not approach questions regarding website owners' discernment of user intent or optimization based on that discernment.

Studies of a more behavioral nature have addressed questions more similar to ours. In one of the first analyses of search engine query data, Silverstein et al. examined queries generally (i.e., not specifically at sponsored search queries) over six weeks of data from the AltaVista search engine query log and found that search engine users differed significantly in their query strategy from users of other data retrieval services and tended to use short queries, the majority using two terms or fewer (1999). This gave rise to several other log-based, query-related studies. For instance, log query data has been used to develop subject-based categorizations (Ross & Wolfram, 2000), determine that a small number of terms are used with high frequency and a large number with a low frequency (Spink, Wolfram, & Jansen, 2001), reflecting a long-tail distribution (Anderson, 2006), and to identify differences in query behavior among users of different search engines (Jansen, Booth, & Spink, 2008).

Research has also considered the effects of user-level variables on search behavior. In this vein, research has found significant differences in search behavior based on gender, cognitive complexity, and cognitive style (Ford, Miller, & Moss, 2005) and based on psychological traits such as extraversion and openness to experience (Heinström, 2003). It has also been found that these individual differences result in distinct search strategies depending on the complexity of the searcher's objective (Ford et al., 2005).

While similar to our search topic in that these research avenues focus on the behavior of the individual, however, they still do not address the idea of determining user intent. Further, to our knowledge, our study represents the first that considers user query-click dyads over time and thus looks at the search session as the unit of analysis and also rests upon a novel, very large data set based on one company's actual sponsored search campaign results.

3.2.2.2. The Search Session

For a website-owning advertiser trying to understand the user's intent, it becomes important to identify the correct unit of analysis. To examine a single search query limits the potential information available; the sequence or collection of search queries made by a user may be both attainable and useful in predicting the user's intent. On the other hand, attempting to view all of a given user's queries over an extended period of time may add considerable noise if the user searches with different intents across that period. To address these concerns, we introduce the idea of the *search session*, a sequence of search query-click dyads related to a single task and conducted by a user over a confined period of time; operationalization of this will be discussed later in this document.

This concept of multiple searches being part of the same overall session has been intimated by earlier research. Using log files, Rutz and Bucklin researched behavior within search engines and found evidence of a potential "spillover effect", in which users who begin with one category of search terms move to other categories, for example starting with a generic term (e.g., "sports car") and moving to a brand-specific term (e.g., "Honda S2000"). Other research has also noted evolution in users' search queries (e.g., Blackwell, Miniard, & Engel, 2006; Browne et al., 2007). These studies, however, have not investigated the ability of accumulated search queries to predict a user's intent once the user arrives at a given website.

3.2.3. Information Search and User Intent

To predict intent, we need to identify the range of intents that may be predicted. *Information search* "is a process in which a person seeks data or knowledge about a problem, situation, or artifact" (Browne et al., 2007, p. 91). Thus, search engine queries — and a great deal of all user activity on the Internet — are acts of information search.

For website owners to identify potential user intents, we thus reference the literature on information search, in particular to streams regarding the purchase funnel and decision-making.

The purchase funnel is a framework studied primarily in the marketing discipline; its intent is to understand stages in a consumer's purchasing process. This pursuit has engendered a number of competing, though not necessarily contradictory, conceptualizations with various numbers of stages. In one of the earliest of these, Lewis viewed the consumer purchase process as consisting of four stages: attention (the consumer becomes aware of a product), interest (the consumer acquires an initial interest), desire (the consumer wants the product), and action (the consumer purchases the product) (1903). These stages are funnel-like in that the consumer initially considers a large number of possibilities, then, over time, narrows these and eventually selects and purchases the one determined to best suit the individual. The funnel itself implies that there should be optimal behaviors for firms in guiding consumers in each stage of the funnel.

While later conceptualizations have included somewhat different stages, they have generally followed a similar, funnel-like progression. Lavidge and Steiner suggest a process comprised of six stages: awareness, knowledge, liking, preference, conviction, and purchase (1961). O'Brien developed another similar model, consisting of only four stages (awareness, attitude, intention, purchase) (1971). Kotler introduced a process consisting of five stages (need arousal, information search, alternative evaluation, purchase decision, post-purchase behavior), with the inclusion of post-purchase behavior novel among the frameworks (1997). Finally, several researchers have simplified the funnel to only two stages, a narrowing stage in which consumers consider a number of possibilities, and a purchase stage (e.g., Andrews & Srinivasan, 1995; Gensch, 1987; Gilbride & Allenby, 2004).

We also find that conceptualizations similar to these have been developed within the broader context of decision making and problem solving. Simon described the decision-making process in three phases: intelligence, design, and choice (1960). Here, the intelligence phase refers to problem recognition and intelligence gathering, the design phase to the structuring of the problem and development of criteria to solve it, and the choice phase the selection of the alternative deemed most fitting. Evidence of the existence of these phases within the Web context has been established (Kohli et al., 2004). In his later work, Simon described problem solving as being like a "search through a maze of possibilities" and states that "successful problem solving involves searching the maze selectively and reducing it to manageable portions" (1981, p. 54). Similar to the purchase funnel, then, in Simon's view problem solving requires a sort of winnowing of possibilities and actions based on the best identified alternatives.

An additional perspective on problem solving stages involves the idea of *task structure*, or the "degree to which the necessary inputs, operations on those inputs, and outputs are known and recognizable to the decision maker" (Browne et al., 2007, p. 92). From a task structure perspective, tasks with higher structure are less complex and vice versa (Byström & Järvelin, 1995). Browne et al. analyzed the decision-making process using two forms of task structure, low and high (2007). Tasks with low structure are construed as potentially having multiple possible answers and a less clear approach to determining what those answers may be, while tasks with high structure tend to have single answers with a clear approach to resolution (D. J. Campbell, 1988). From this standpoint, the task structure itself may indicate the individual's stage in the decision-making process.

The overlap among these views on information search processes confirms an underlying premise. Learning about a product, searching a possibility maze to solve a problem, and considering a low structure task all require broad information and cognitive abstraction. Meanwhile, making a product purchase, identifying and executing a decision, and completing a high structure task reflect an already narrowed set of possibilities. Thus, when predicting the intent of users arriving on websites, we propose that predicting this intent based on funnel location, problem solving stage, or task structure should all reflect similar relationships.

Website owners would, then, expect individuals at the top of the purchase funnel, early in the problem-solving process, or completing an unstructured task to behave in similar ways. These users need to identify broad possibility sets from which to narrow down possible products or answers as well as to attain the information needed to structure their decision-making. Therefore, the website owner should expect that the search session of a user at the top of the funnel (i.e., in the information-gathering stage) would be more likely to use a broader set of search terms, while someone closer to the purchase stage would use a narrower set of terms. Indeed, in the offline context, Biehal and Chakravarti found that consumers searching for information about a product purchase will, as they progress down the funnel, begin to search for information about ever more specific brands and products (1983). Similarly, we expect that those who have a well-structured problem or who have already identified the product to purchase will search using more specific terms (e.g., make and model of a product, a narrowly detailed question).

These observations bring us to our two basic hypotheses:

> *Hypothesis 1: Users searching broadly will be more likely to be at the top of the funnel (i.e., gathering information) and those searching less broadly will be more likely to be at the bottom of the funnel (i.e., completing purchase).*

> *Hypothesis 2: Users searching more specifically will be more likely to be at the bottom of the funnel (i.e., completing purchase), while those searching less specifically will be more likely to be at the top (i.e., gathering information).*

Beyond this, it has been found that individuals' searches can vary based on goal types and complexity, as well as their cognitive style (Ford, Miller, & Moss, 2001; Ford et al., 2005). As a result, we expect that user behavior should be categorizable as a result of this diversity of complexity and, further, that these categories may be useful in predicting user intent upon arrival on the website. These concepts will be investigated in our data analysis.

3.3. Data

3.3.1. Data Source

To perform our investigation, data were acquired from a leading online travel agency in Asia (XYZ Travel hereafter), which conducts sponsored search advertising. Data acquired were based on advertisements placed through a single platform. From these advertising activities, XYZ Travel has accumulated individual user search and click data, which can be matched to online sales results for that user. The raw data contains information about each user's sponsored search advertisement click that led the user to the XYZ Travel website. This includes the search query entered, the time of the query, the user's IP address, and user purchase data (i.e., whether the user purchased or not)[6]. The dataset contains a total of 2,399,391 raw search cases collected during a one-year span; 172,671 cases were generated by users who made purchases, while 2,226,720 were generated by users who did not. The total number of unique keywords used was 11,221. The dataset also includes repeat searches from the same user; retaining all user search cases was necessary for session identification as described in the next section.

A data set of this size is expected to have a wide range of search types that would enable significant variance in search session query breadth. Our underlying assumption is that any search that resulted in a click on an XYZ Travel advertisement was performed to find information to help the user make a travel-related decision (i.e., solve a travel-related problem).

Each query in our data set can be sorted by IP address and time, yielding sequences of queries that can be divided into search sessions. We use IP address as a surrogate for a discrete user; in other words, we assume that two searches made from the same IP address were conducted by the same individual. Examples of sorted keywords by user are shown in Table 3.1. After sorting, the number of unique users

[6] Note that search queries were only captured when they resulted in clicks on an advertiser's ad. Queries entered that did not result in such clicks were not captured.

(based on IP address) in the data set is 1,176,115, of whom 80,840 made purchases. The mean number of searches per search session was 2.13 for purchasing customers and 2.01 for non-purchasing customers.

IP Address	n of Queries	Query 1	Query 2	Query 3	…	Query 15
158.xxx.xxx.xxx	2	Airplane Ticket	Price Comparison			
210.xxx.xxx.xxx	15	Singapore Airtel	Guam Airtel	Airline Reservation	…	Hong Kong Travel
…	…	…	…	…	…	…
165.xxx.xxx.xxx	1	Tokyo Travel				

Table 3.1. Examples of search query sequences.

3.3.2. Session Identification

Data were collected over the course of 12 months. Given this extended time period, it is assumed that some queries from the same IP address likely consisted of multiple search sessions; that is, that the same user pursued different search objectives over the course of those 12 months. Given our selection of the search session as our unit of analysis, operationalizing this concept is of great importance. A study by Göker and He established 11 minutes as the optimal criterion for separating on-site search sessions using server log data (2000). By this approach, if the same user's search sequence experiences a break of 11 minutes or more (i.e., at least 11 minutes transpire between searches), then the next query should represent a new search session.

Given that this criterion was established based on on-site search, rather than sponsored search queries and clicks, we conducted further analysis to determine whether this would be the best approach for search session identification. Three graduate students in Information Systems were hired as search session coders and a sample of 5,920 queries from 200 randomly selected IP addresses was used. These queries were time-stamped, allowing the coders to see the time between queries. The coders were then asked to separate the sequences of queries into sessions based on the queries' content using a logical rationale. For instance, if three consecutive queries seemed to refer to a trip Hong Kong, then the fourth to a trip to Hawaii, then a session break could have been identified between the third and fourth queries. Further, if a sequence of search terms included an apparently related purchase, then the search session ended at the purchase, with the next query instigating a new search session.

The graduate students' coding was compared against nine different candidate criteria: Göker and He's 11 minutes criterion, a similar criterion using 24 hours instead of 11 minutes, average of all time intervals for the given IP address, median of all intervals for the given IP address, mean absolute deviation (MAD) of all IP address intervals, mean interval plus two standard deviations, mean interval plus three standard deviations, MAD plus two standard deviations, and MAD plus three standard deviations.

Among these criteria, we found that the Göker and He criterion performed most reliably. This 11-minute criterion yielded the same results as those coded by the three coders in 68.9%, 71.0%, and 71.7% of cases respectively. The other candidate criteria performed significantly worse, with the highest reliability among them being 55.4%; see Appendix B, Table B.1 for a comparison of performance for the candidate criteria. We thus identified the 11-minute criterion as the most appropriate for identifying search sessions among the users' query sequences, and therefore used this as the criterion. This coding thus resulted in the identification of 1,823,539 discrete search sessions, of which

17,708 resulted in a purchase. Users who made a purchase had a mean of 3.46 queries per session, while non-purchasers had a mean of 1.10.

3.3.3. Keyword Coding

Coding of keyword data presented one of the biggest challenges of this study. This coding is particularly critical as both search sessions and dimensions of these sessions were coded through this process. Without similar previous studies for guidance, the researchers employed a two-phase approach for keyword coding. In Phase 1, researchers determined a typology for query dimensions (e.g., "location specificity"). Based on the specific criteria and the information captured, each of these dimensions was designated as binary, categorical, or ordinal in scale. In Phase 2, queries were coded by multiple coders, aided by an automated process, according to the criteria developed through Phase 1.

The Phase 1 typology of query dimensions was developed based on a thorough review of random samples of query sequences. After establishing this initial typology, a sample of 40 users each from both purchasing and non-purchasing user groups was coded by the researchers individually by hand to verify the validity and replicability of these categorizations. Discrepancies among coding results were discussed and dimensions refined based on these discussions. A second round of sample coding was then completed in a similar manner; researchers coded sample keywords and any disagreements or possible problems were discussed and the coding scheme modified accordingly. Three rounds of sample coding were carried out, after which the coding scheme was deemed final as agreement had been reached among the researchers using only the criteria set forth in the previous rounds. The dimensions included in the final scheme include: travel location, location specificity, airline, type of activity, purpose of travel, reservation-related terms, recommendation-related terms, price-related terms, advertiser name (i.e., "XYZ Travel"), and otherwise uncaptured specificity. See Table 3.2 for more details.

Dimension	Description
Location Level	0 = N/A, 1 = Region (e.g., Europe), 2 = Country (e.g., Japan), 3 = City (e.g., Los Angeles), 4 = Attraction (e.g., Phuket Beach, Disneyland)
Type of Airline	0 = N/A, 1 = Domestic Airline, 2 = Non-Domestic Airline
Type of Activity	0 = N/A, 1 = Package, 2 = Semi-Packaged ("Free") Travel, 3 = Backpack Travel, 4 = Transportation (e.g., airplane, train), 5 = Lodging Type
Purpose of Travel	0 = N/A, 1 = Business (e.g., conference, exhibition), 2 = Active Leisure (e.g., ski, golf), 3 = Inactive Leisure (e.g., beach, cruise)
Schedule	0 = N/A, 1 = Includes a schedule-related term (e.g., "itinerary")
Reservation	0 = N/A, 1 = Includes a reservation-related term (e.g., "booking")
Purchase	0 = N/A, 1 = Includes a purchase-related term (e.g., "buy")
Price-Related	0 = N/A, 1 = Includes a weaker price-related term (e.g., "price check"), 2 = includes a stronger price-related term (e.g., "discount", "lowest price")
Recommendation	0 = N/A, 1 = Includes a recommendation-related term (e.g., "popular")
Advertiser Name	0 = N/A, 1 = Includes "XYZ Travel"
Other Specificity	0 = N/A, 1 = Includes a specific term not captured in the above dimensions

Table 3.2. Query dimension coding scheme.

Phase two consisted of both a screening stage and a coding stage. In the screening stage, seven undergraduate research assistants were recruited as coders and asked to screen for and identify specific destinations, airlines, and other terms frequently appearing in the search queries. Once completed, these terms were used in the coding stage.

In the coding stage, ten coders coded queries using Web-based software developed by the researchers specifically for this purpose. A portion of the coding was completed automatically by a text analysis program, which identified dimensional values for the most commonly used query terms. For instance, there was a relatively small number of popular destination terms commonly used among the queries, such as

"New York", "Shanghai", and "Hawaii". These were identified by the purpose-built software and coded appropriately. Other common words and phrases were likewise identified and coded. Further, for more accurate coding and to reduce subjective bias that might arise, each query requiring manual coding was coded by two independent, randomly assigned coders, thus following examples set in prior studies (e.g., Boudreau, Gefen, & Straub, 2001; Shrivastava, 1987).

3.3.4. Specificity and Breadth

3.3.4.1. Specificity Measures: Location Depth and Other Depth

Testing the effects of specificity of search session queries requires us to define specificity within this data set using the dimensions coded. Here we identified location depth as one measure of specificity and other depth as another. Location depth is determined here based on the location level coded as described above (0 = no location, 1 = region, 2 = country, 3 = city, 4 = attraction) of each query. The location depth of a search session was then measured as the mean of location depths of all queries within that session.

The "other depth" measure of specificity captures the extent of details of non-location query terms within a search session. Several of the variables determined in Phase 1 and included in Table 3.2 are associated with search specificity: type of airline, type of activity, purpose of travel, schedule, reservation, purchase, price-related, recommendation, and advertiser name. These variables were combined into a new variable, "other depth", which was applied to each query and then each search session. Specifically, the number of these variables having a non-zero value for each query was counted as the other depth for the query. For instance, a query of "buy asiana airlines tickets at xyz travel" would have received an other depth score of three, one for a purchase-related term ("buy"), one for an airline named ("asiana airlines"), and one for the name of the advertiser ("xyz travel"). The other depth value for the entire session was then calculated by averaging the values of all queries included in the session.

3.3.4.2. Breadth Measurement

Breadth measures the diversity of terms used within a query; as a search session increases in the number of dimensions referenced, the breadth increases. For example, a session may include a query for "Tokyo hotel" followed by a search for "Tokyo subway". In this case, the second query increases the session's breadth since it adds an element ("subway") belonging to the transportation categorization that was not part of the first query.

As with specificity, we measured breadth in terms of both location and other. Location breadth measures the diversity of locations in a session; the more locations referenced in a session, the larger the location breadth. For instance, a search session that included queries of "Los Angeles hotel", "Las Vegas hotel", and "Hawaii hotel" would have a location breadth of three. Note that nested terms were not considered to add breadth so that a session that began with the term "Japan tours" would not see its location breadth increased by the inclusion of the query "Osaka tours", since Osaka is nested within Japan. For other breadth of a session, the number of specificity-related dimensions (type of airline, type of activity, purpose of travel, schedule, reservation, purchase, price-related terms, recommendation, and advertiser name) in the query that have non-zero values were added together. A session's location breadth and other breadth were both determined by taking mean values across all queries.

3.4. Analysis

3.4.1. Breadth and Specificity

In our basic hypotheses, H1 and H2, we expect that user intent can be inferred from their on-site behavior and, specifically, that attributes of search sessions can be used to predict whether a user will make a purchase. Since the purchase is a binary variable (either the search session resulted in a purchase or it did not), we conducted a logistic regression analysis to test H1 and H2. As earlier research has found that time spent searching affected user behavior in the broader internet context (Johnson, Moe, Fader, Bellman, & Lohse, 2004), we also control for total session time in this model[7]. In order to check for multi-collinearity, an OLS regression was carried out prior to the logit analysis (Menard, 2002). The highest variable inflation factor (VIF) found was 2.47, indicating that no serious multi-collinearity is present. Descriptive statistics and a correlation matrix for referenced variables can be found in Appendix B, Tables B.2 and B.3 respectively. The logit regression results are summarized in Table 3.3 below and prediction accuracy in Table 3.4.

[7] Past research has also found number of queries to significantly affect user behavior, however, as this is highly correlated with session duration in our dataset and is nearly constant within the full data set, we have omitted it in our model.

Variable	B	SE	p	Exp(B)	Supported?
Constant	-5.973				
Location Breadth	-3.958	0.195	< 0.001	0.019	H1 Supported
Other Breadth	-3.871	0.121	< 0.001	0.021	H1 Supported
Airline Name	0.387	0.076	< 0.001	1.473	H2 Supported
Brand Name	0.641	0.026	< 0.001	1.898	H2 Supported
Location Depth	-0.008	0.007	0.261	0.992	H2 Not Supported
Other Depth	0.272	0.014	< 0.001	1.312	H2 Supported
Session Duration	0.457	0.002	< 0.001	1.579	

Table 3.3. Breadth and specificity test results, full data set. Nagelkerke $R^2 = 0.369$, -2LL = 128,382.71.

	Predicted		
	No	**Yes**	**% Correct**
Observed No	1,803,558	2,273	99.87%
Observed Yes	14,682	3,026	17.09%
Overall			99.07%

Table 3.4. Prediction accuracy, full data set.

Using the full data set, we find that broader searches (i.e., those less-focused on narrow topics) significantly relate to a lower probability of conversion (hence the log odds ratios below 1), both in terms of location breadth and other breadth[8]. We also find support for the relationship between specificity and conversion. Specificity, measured in terms of other depth, was significantly related to conversion. Further, both brand specificity and airline name specificity significantly contributed to higher conversion. We note, however, no significant effect based on location depth. However, the preponderance of evidence suggests that users searching less broadly as well as with greater specificity can be expected to have a stronger intent to purchase. The model accurately predicts purchase 99.07% of the time, versus 98.89% accuracy if predicting “no purchase” for each session.

This represents a significant improvement (0.18 in absolute points compared to a possible improvement of only 1.11 absolute points before 100% accuracy is reached;

[8] We note here that both measures of breadth are near constants and both have mean values of 0.01. Results of a logistic regression with these two variables omitted can be found in Appendix B, Table B.4.

accuracy is this improved 16.2%). However, we note that the accuracy of predicting purchasers, the more interesting of the two groups for advertisers seeking to attract only those searchers lower in the funnel, is poor. This low prediction accuracy is not unexpected; in data sets with binary dependent variables where one of the variables is many times more frequently represented than the other, logistic regression underestimates the likelihood of the "rare event", the purchase in this case (King & Zeng, 2001). To address this issue, we created a subset of the data that included 5,000 randomly selected purchasers and 5,000 randomly selected non-purchasers and again estimated a logistic regression (King & Zeng, 2001)[9]. This approach results in models with less accurate intercept coefficients (compared to the coefficient identified through the full data set model), but with slope coefficients for the independent variables that remain unbiased (Allison, 2012). Results of this estimation can be seen in Tables 3.5 and 3.6 below.

Variable	B	SE	p	Exp(B)	Supported?
Constant	-9.293				
Location Breadth	-8.685	0.730	< 0.001	< 0.001	H1 Supported
Other Breadth	-6.246	0.978	< 0.001	< 0.001	H1 Supported
Airline Name	0.109	0.242	0.654	1.115	H2 Not Supported
Brand Name	0.673	0.175	< 0.001	1.961	H2 Supported
Location Depth	0.032	0.040	0.421	1.033	H2 Not Supported
Other Depth	0.143	0.077	0.064	1.154	H2 Weak Support
Session Duration	0.240	0.021	< 0.001	1.271	

Table 3.5. Breadth and specificity test results, evenly distributed data subset. Nagelkerke $R^2 = 0.894$, -2LL = 2,760.120.

[9] We also conducted a rare event logistic analysis using the full data set. Results of this regression were similar to that of the initial logistic regression also using the full data set. See Appendix B, Table B.5 for results of the rare events logistic regression.

	Predicted		
	No	**Yes**	**% Correct**
Observed No	4,633	367	92.66%
Observed Yes	4	4,996	99.92%
Overall			96.29%

Table 3.6. Prediction accuracy, evenly distributed data subset.

Adjusting for rare events in this manner produces a few interesting differences in our results. First, the pseudo-R^2 (Nagelkerke) and -2 log likelihood values suggest that this may be a better model, although given the difference in sample size and DV distribution, quality of fit is difficult to determine. Second, we see weakened support for one of our two hypotheses. While H1 regarding the effects of breadth is still supported for both measures of breadth, H2, regarding specificity of search terms, received less support than in the previous model. Finally, while the overall prediction rate is lower in this model than in the previous (96.29% vs. 99.13%), we note that this represents a 46.29 percentage points improvement over predicting either one result or the other in all cases (which would yield only 50% accuracy). Thus, this model increases prediction accuracy by 92.58%.

One of the strengths of this data set is that it contains multiple queries from the same user within the timeframe necessary to be considered a "search session" based on our criteria explained earlier. We find, however, that the mean number of queries per session is 1.01 — in other words, most sessions consist of only one query. What happens, however, when we consider the sessions with multiple queries? Do our hypotheses still hold? To examine this, we divided our data set appropriately and conducted a logistic regression analysis of multiple-query search sessions[10]. Descriptive statistics for this sub-sample (n = 133,372) can be found in Appendix B, Table B.6. The findings are in Tables 3.7 and 3.8 below.

Variable	B	SE	p	Exp(B)	Supported?
Constant	-3.506				
Location Breadth	-3.151	0.195	< 0.001	0.043	H1 Supported
Other Breadth	-2.933	0.094	< 0.001	0.053	H1 Supported
Airline Name	0.043	0.063	0.502	1.044	H2 Not Supported
Brand Name	0.426	0.025	< 0.001	1.531	H2 Supported
Location Depth	0.014	0.008	0.091	1.014	H2 Weak Support
Other Depth	0.302	0.019	< 0.001	1.352	H2 Supported
Session Duration	0.201	0.002	< 0.001	1.222	

Table 3.7. Breadth and specificity test results, multiple query sessions only. Nagelkerke R^2 = 0.215, -2LL = 76,951.710.

	Predicted		
	No	**Yes**	**% Correct**
Observed No	117,877	916	99.23%
Observed Yes	12,927	1,652	11.33%
Overall	130,804	2,568	89.62%

Table 3.8. Prediction accuracy, multiple query sessions only.

We see, then, increased support for the importance of location depth. Airline name continues to be a weak predictor of funnel location (i.e., purchase behavior). However, for website owners who track users over multiple query-click dyads within a single search session, we continue to see the ability of both breadth and specificity measures to predict funnel location.

From this point, we also explored whether purchasers who began by looking for specific terms would spend less time searching, thus giving us further evidence that these users are further down the funnel and pursuing a more transactional intent. To test this hypothesis, we performed a regression analysis with the following results.

[10] While comparing multiple-query sessions to single-query sessions would appear interesting, of the 1,652,531 single-query search sessions, only seven resulted in a purchase. See Appendix B, Table B.7 for descriptive statistics of the data set sub-sample including only single-query sessions.

Variable	B	SE	Std. B	p
Constant	2.680			
Location Breadth	2.805	0.989	0.059	0.005**
Other Breadth	3.902	0.454	0.105	< 0.001**
Start w/ Airline Name	-0.192	0.280	0.151	0.492
Start w/ Brand Name	-0.414	0.107	0.193	< 0.001**
Location Depth	-0.642	0.037		< 0.001**
Other Depth	0.171	0.065		0.045*
Number of Queries	1.526	0.021		< 0.001**

Table 3.9. Starting specificity test results. $R^2 = 0.595$.

Here, we see mixed support for this idea. Users who begin their search sessions with the brand name ("XYZ Travel") indeed spend significantly less time searching. However, for those who begin with a specific airline name, we do not find a significant relationship (although the sign is in the predicted direction). This may be caused by a number of factors, including an ineffective link (i.e., the ad's link does not send users to a page that includes expected information) or the possibility that those searching for airline name may still be price shopping and, thus, may still be higher in the funnel and still in need of more information and related longer search sessions. It is also possible that many users may have frequent flier accounts on particular airlines and thus always begin their travel-related browsing in that way, regardless of funnel location.

3.4.2. Search Session Classification

Finally, we analyzed the data to determine whether we could identify categorizable search session behaviors that provide meaningful insight into user intent, again focusing on funnel location. Behaviors included our previously discussed variables: depth (location and other), breadth (location and other), and specificity (combined airline name and brand name). For the specificity, we used the sum of our airline name and brand name variables, each of which was a binary variable. However, the resulting specificity variable had a range of only 0 to 1, not 0 to 2. We also included session duration and number of queries. Because of the presence of the binary specificity variable, we used two-step clustering analysis, which identified six discernable search patterns (see Table 3.10 below).

Variable	1.	2.	3.	4.	5.	6.
Location Breadth	0.00	0.42	0.00	0.00	0.00	0.00
Other Breadth	0.68	1.34	0.00	0.00	0.00	0.00
Location Depth	0.00	0.65	0.42	2.86	0.06	0.31
Other Depth	1.42	1.03	2.00	1.00	1.00	1.47
Specificity	0.00	0.00	0.00	0.00	0.00	1.00
Query Count	2.94	2.21	1.00	1.01	1.00	1.05
Session Duration	7.55	4.00	0.00	0.00	0.00	0.01
n	*98,703*	*22,241*	*421,019*	*327,901*	*300,626*	*147,399*
% of Sample	*7.5%*	*1.7%*	*31.9%*	*24.9%*	*22.8%*	*11.2%*
Purchase %	*11.42%*	*13.75%*	*< 0.01%*	*0.03%*	*< 0.01%*	*0.09%*

Table 3.10. Classification of search session behavior.

We note here that, based on purchase behavior, two of these groups are clearly lower in the funnel than the other four. These lower funnel groups, 1 and 2, appear to have a significant separation from the higher-funnel groups in terms of a few key criteria: query count, session duration, and breadth (although this is likely a function of these groups' sessions including multiple queries). Between the two lower-funnel groups, we note that group 1 had the longest mean query duration and the highest mean session count of any group, while group 2 had the broadest terms. Thus, those sessions reflecting the lower funnel locations are those that include multiple queries and have longer durations.

We then conducted an analysis of variance to determine whether these groups will be significantly different in their ability to predict funnel location (by predicting purchase behavior). There was, in fact, a significant effect for group membership at the $p < 0.05$ level for the six conditions, $F(5,1317886) = 31681.179$, $p < 0.001$. Post hoc comparisons using the Tukey HSD test indicate that the mean of purchases for Group 1 ($\bar{x}$ = 0.14, s = 0.344) was significantly different from Groups 2 through 6. The mean of purchases for Group 2 ($\bar{x}$ = 0.11, s = 0.318) was significantly different from Groups 3 through 6. See Table 3.11 for descriptive statistics of the analysis; details of post hoc comparisons can be found in Table 3.12.

Group	n	Mean	S.D.	S.E.
1	98,703	0.11	0.318	0.001
2	22,241	0.14	0.344	0.002
3	421,019	< 0.01	0.002	< 0.001
4	327,904	< 0.01	0.018	< 0.001
5	300,626	< 0.01	0.002	< 0.001
6	147,399	< 0.01	0.030	< 0.001
Total	1,317,892	0.01	0.105	< 0.001

Table 3.11. Descriptive statistics for between-subjects ANOVA.

	1.	**2.**	**3.**	**4.**	**5.**	**6.**
Group 1	–	-0.023**	0.114**	0.114**	0.114**	0.114**
Group 2	0.023**	–	0.137**	0.137**	0.137**	0.137**
Group 3	-0.114**	-0.137**	–	< 0.001	< 0.001	-0.001*
Group 4	-0.114**	-0.137**	< 0.001	–	< 0.001	-0.001
Group 5	-0.114**	-0.137**	< 0.001	< 0.001	–	-0.001*
Group 6	-0.114**	-0.137**	0.001*	0.001	0.001*	–

Table 3.12. Group comparisons. Mean difference shown; * = significant at α = 0.05, ** = significant at α = 0.01.

Both groups 1 and 2 are significantly more likely to predict purchase than groups 3 through 6, suggesting that website owners might reasonably and algorithmically determine user intent and, based on that intent, deliver optimized content.

3.5. Discussion

3.5.1. Results and Implications

As demonstrated in our basic hypotheses (H1 and H2), user intent can indeed be inferred based on characteristics of the user's arrival on a website. Those who arrive having searched for less broad and more specific terms can generally be expected to more likely be pursuing a low-funnel, high-structure task such as making a purchase, while those who have searched for broader and less specific terms can more likely be expected to be pursuing a low-structure task.

Interestingly, there are nuances to this. As we saw, not every measure of specificity was significant in predicting funnel location. In the full data set, Location Depth was insignificant, but became weakly significant in the multiple-query sessions sample. Further, while Airline Name was significant in the full data sample, it was insignificant in both the 50/50 (purchase/no-purchase) split sample as well as the multiple-query sessions sample. From this, we suggest that while specificity matters, not all types of specificity matter in all cases. From a website owner's perspective, this implies an important need to test around the various kinds of specificity to determine which are important in predicting low-funnel locations for website visitors. In the case of the fictitious projector company mentioned at the outset of this study, they may find that increased specificity in terms of the company's model numbers is meaningful, but specificity in terms of product features is not.

On the other hand, we found that both measures of breadth signal a user who is further up in the purchase cycle. This was found to be the case in all subsets of the data. As users' query-click dyads were broader and less focused, entailing multiple keyword types (e.g., a destination and a type of transportation, a lodging type and an activity), the more likely they were to be further up the funnel. Given that we found this to be true for both forms of breadth and in all cases, we suggest that practitioners may be able to assume that broader queries, regardless of the variety of keyword categories from which they draw, can generally predict higher-funnel user intent.

Further, our user classifications show that user sessions can be predicted into categories that significantly relate to probability of purchase and, thus, to a location lower in the funnel. The criteria from such a categorization exercise could be used to tailor website experiences to the predicted user intent based on search session query information delivered electronically to the website for the arriving visitor. For instance, those users whose intent is algorithmically determined to be more probably low-funnel might be directed to a landing page that is more geared toward sales messaging in order to better facilitate the completion of a transaction. Those users who are higher in the funnel, on the other hand, may have a better experience if website messaging and navigation facilitates — via strong information scent (Card et al., 2001), for instance — access to information that can guide such users toward lower levels of the funnel.

Further, as mentioned, sponsored search campaigns represent large marketing investments. Websites themselves can be very resource-intense to design, build, and maintain. Understanding the user's intent upon arrival can help website owners optimize that user's experience, for instance by focusing on helping the user complete their high-structure task efficiently or in presenting more information in an accessible way to help users with a low-structure task to begin the information winnowing process. This study represents a first attempt at discerning user intent based on information passed to the website upon arrival. It may be that other arrival types (e.g., those coming from a search engine, those coming via direct URL entry, etc.) can further strengthen the ability of website-owning firms to infer intent.

We also find that the vast majority of searches do not result in low-funnel activities. This preponderance brings to a head the question of the ability of companies to place a value on such searches in order to place appropriate bids through sponsored search and to evaluate expenditures against user behavior. Further, while these users may not be in the low-funnel purchasing mode, it may be useful to understand how far up-funnel they might be. If a firm understands that a user is still determining whether a product is needed, it may be beneficial to communicate with that user in a different way than would be most appropriate for a user who has already decided the product is needed and is now comparing models against one another. Diving even more granularly into specific qualities of search session queries may cast further light on the intent of users and, thus, enable website owners to understand more exact up-funnel positions of website visitors. This may be a useful avenue for future research.

3.5.2. Limitations

We acknowledge there are limitations to our findings. First, while the paper builds on theory in several respects (our conceptualizations of user intent and the user search session, for instance), the analysis of the data are largely atheoretical. We follow in this a body of IS research that has followed this approach, using the data itself to inform research (Johnson et al., 2004). While this approach enabled us to make discoveries regarding the usefulness of breadth and specificity, we acknowledge that without the support of a greater body of established theory, generalizability of our findings may not be reasonable without further research to confirm our findings.

Relatedly, our data set comes from a specific advertiser (an Asian travel agency). While we do not expect that customers from other regions of the world or customers of other product types would differ significantly in their behavior and would thus invalidate our findings with regard to predicting user intent, it may be useful to replicate this study in other contexts to verify that this is the case. Further, while our dimensions of specificity included details such as “airline name” and depth was specific to “location”, we feel that these concepts would have cognates in other industries. We note that identifying such cognates would be an important step in applying our findings to other contexts and data sets.

We also note that this study uses as its unit of analysis not the individual, but rather the user search session. This decision was made so as to isolate the user’s information seeking and purchase process for the purchase of a single item. Looking at a user’s activities among various discrete sessions would potentially capture multiple buying processes (e.g., Process A to book a trip to Fiji, Process B to find airfares for relatives coming from California, Process C to book an all-inclusive leisure tour to Hong Kong). However, by taking this approach and basing sessions as we did on time gaps between searches, we may have inadvertently separated individual search processes. While this may have thereby reduced predictive power of the analysis, we feel this was nevertheless a conservative approach since it would have only served to decrease the likelihood of finding significant results. Future researchers may wish to explore additional alternative means of identifying discrete sessions in order to increase the predictive power.

Similarly, in our coding of data, we considered unique IP addresses to indicate unique individuals. This may not always be the case, for instance if the same IP address is shared by multiple work associates at the same office or multiple members of the same household. However, again, we believe that this represents taking a conservative approach as this potential noise would have only reduced predictive power. Again, future researchers may wish to find approaches to better isolate individual users without making the assumption regarding IP address.

Predictions within this analysis make an assumption that all users coming by way of a search engine are seeking to complete a task. It is possible that some intents cannot be so construed. While one may argue that even a hedonic purpose remains a "task" when it involves using an online search engine, it may be useful for future researchers to study hedonic searchers in isolation as has been done in other areas of the information systems literature.

Finally, while this research remains an important first step in predicting user intent upon arrival on a website, it does so only within the specific context of a user arriving via search engine. The field would benefit from research and discoveries regarding arrival through other means and mechanisms.

3.5.3. Conclusion

This research contributes to the literature by introducing a typology for search query keywords as well as the concept of the user search session in the sponsored search context. We further introduced the use of the purchase funnel as a means of describing user intent and showed the parallels between this approach and the concepts of problem solving (Simon, 1960) and task structure (D. J. Campbell, 1988). These can be particularly helpful in fostering further research into the meanings and implications of user search terms as well as enabling future researchers to isolate purchase and task processes.

Further, we showed how a dataset can be used to track user search processes across multiple searches. From this, we showed not only that search term breadth and specificity matter, but also contributed the idea that different types of specificity may matter differently or not at all. We also made a contribution by showing that, indeed, a user's intent can be inferred based on information related to the user's means of arrival. This possibility of inferring user intent upon arrival can be of great use for website owners seeking to improve the success of their websites by adjusting user experience based on such inferred intent.

Digital Borders: Location Perception and Success Attribution in the Web Environment

4.1. Introduction

Where are you? The question is not as simple as it may at first seem. Complications arise, for instance, from the multitude of possible answers: in an armchair, in the living room, in my house, in North America, on Earth. Answering the question increases in complexity when travel is involved and as an individual moves among locations. Driving down the freeway, a person transits among various neighborhoods, cities, counties, and regions, and may only become aware of which locality is being entered when a sign at the side of the road proclaims the transition loudly enough. Even then, the motorist may be aware of being, for instance, in the city of Pittsburgh without knowing the name of the exact neighborhood through which the car is traveling. Going another 30,000 feet skyward, an airline passenger likely has very little concept of which borders are being crossed and when. This complexity and ambiguity can have some real repercussions. Laws may vary among cities and counties, a license in one state may not be valid in another, and the beautiful countryside seen from the airplane window will be difficult to visit later on without knowing what it might be called or within which exact borders it is located.

Perceiving locations and borders online presents even more complexity. Traveling between two websites takes neither significant time nor effort; Web users are constantly moving from one page to the next, sometimes resulting in their departing a site. Users may be aware of a change in location, but may not comprehend where they have gone. Further, while the inherent hierarchy within a physical geography is generally clear (e.g., a street address within a city within a county within a state), the online hierarchy may be less clear and therefore add to confusion about location. If a user follows a link on a search engine, is the resulting page seen as falling under that search engine in the hierarchy? When, for instance, a user researches a news story through Google News and finds the sought-after information at the *New York Times*, did the user just have a Google News experience, or was it a *New York Times* experience?

This potential for online location confusion matters. Site owners need users to understand where they are in order to generate loyalty and build an audience. In particular, in recent years content-based commercial websites (e.g., CNN.com, Wired) have seen the number of users who come to their homepage deteriorate. According to a *New York Times* internal report, the newspaper's home page traffic decreased about 50% from 2011-2013 as more and more users accessed parts of the site through links from other websites (e.g., search engines, blogs) (Tanzer, 2014). The same shifting of traffic sources has been reported by other content-based sites as well (Thompson, 2014). Users arriving on pages in "the middle" of a website, rather than through the "front door" of the homepage, is roughly equivalent to a traveler materializing in the lobby of San Francisco's Transamerica Pyramid Center, rather than entering the city by way of the

Bay Bridge; in the former case, the traveler has few cues indicating the location of materialization, while in the latter the crossing of the San Francisco bay and signage at the city limits give more obvious clues. For content websites, if users do not know where they are, they are less likely to generate a preference for a given online location.

Along these same lines, consider the idea of a site owner inferring success through website usage metrics. However, if the users visiting the site failed to recognize the identity of the site with which they interacted, does it matter how many pages they viewed there or how many minutes they spent “engaging” with the site’s sponsor? Further, users may attribute success (or failure) to the location that they perceive themselves as using, rather than the one they actually use. This could result in a company doing a good job and not getting the appropriate credit or, perhaps worse, receiving unfair blame for something out of its control.

From the user’s standpoint, site policies are an important protection against potential privacy violations. However, without knowing under exactly whose policy they fall, users may be confused about whom to hold responsible for privacy breaches or may not understand which specific safeguards are in place. The potential for such confusion has proven a particular issue for Google, against whom European regulators have threatened legal action as a result of its combining privacy policies across its owned sites (e.g., the Google search engine, YouTube, Google+, etc.). Regulators there feel that Google’s perception of itself as a single entity may not match users’ perceptions of location and ownership, thus potentially raising privacy issues for users (Davenport, 2013).

This ambiguity around website borders also affects research. Investigations into user interactions with online platforms have tended to view website borders as clear and known. Websites have generally been treated as discretely perceived entities with which users interact intentionally while recognizing their formal borders. This approach has been useful in exploring a number of behavioral concepts including flow (Koufaris, 2002), cognitive absorption (Dunn et al., 2014; Lowry et al., 2013), telepresence (Coyle & Thorson, 2001), interactivity (Albert et al., 2004; Coyle & Thorson, 2001; Dunn et al., 2014; Y. Liu, 2003; Lowry et al., 2009; Palmer, 2002; Song & Zinkhan, 2008), usability (Agarwal & Venkatesh, 2002; Lowry et al., 2008; Venkatesh & Ramesh, 2006), vividness (Coyle & Thorson, 2001), delay (Galletta et al., 2006; Palmer, 2002), annoyance (McCoy et al., 2007), trust and distrust (Gefen et al., 2003b; Lowry et al., 2008; McKnight et al., 2002; Pavlou & Fygenson, 2006), privacy (Hui et al., 2007; Lowry, Cao, & Everard, 2011), design (Zhenhui Jiang & Benbasat, 2007; Nadkarni & Gupta, 2007), and user acceptance (Gefen et al., 2003b; Pavlou & Fygenson, 2006). While clearly useful, in limiting itself to discrete content sites, this approach ignores the idea of border ambiguity and border perception. In practice, users may not always be aware of having crossed from Site A to Site B. Further, the single-site approach ignores any influence of Site A on Site B and the effects of travel across borders are not recognized.

This study explores this concept of borders among websites. In doing so, we open new avenues of inquiry that may eventually help explain the value among links in different types of digital content chains, the effects of various routes that users may take to arrive at content, the cognitive load of border crossings, and issues of intellectual property rights. To open these avenues, we address key questions regarding the perception of borders within the online environment. Are users aware of crossing borders and, thus, of location? What factors affect this awareness? How do these factors and any resulting awareness impact the users' willingness to attribute credit to the locations involved in a Web-based experience?

In addressing these questions, we first discuss the concept of location within a digital content (website) environment, invoking the theory of Space and Place (e.g., Buttimer, 1976; Mennecke, Triplett, & Hassall, 2011; Relph, 1976; Saunders, Rutkowski, von Genuchten, & Vogel, 2011; Tuan, 1974, 1977), which explains how a location acquires meaning for an individual. We then discuss the existence of online borders, both formal and perceptual. We apply these concepts to the visitation of online locations to derive hypotheses, which are then tested using data collected in an Web-based experiment. Following analysis, the implications and future research opportunities suggested by this study are then discussed.

4.2. Background

4.2.1. Space and Place

Given that existing IS literature has not focused specifically on the subject of borders within the digital content context, we turn to an external literature and the geographic study of physical location. The study of location in the physical world has interested researchers at least as far back as Aristotle, who contemplated its meaning in his *Physics IV* (Morison, 2002). Starting in the 1970s, experiential geographers began formalizing concepts and theory around the idea of "space and place"(e.g., Agnew, 2011; Buttimer, 1976; Gustafson, 2001a; P. Lewis, 1979; Relph, 1976; Tuan, 1974, 1975, 1977, 2001). According to this body of work, *place* is defined as a location to which meaning is attached, while *space* refers to locations without existing meaning (P. Lewis, 1979; Relph, 1976; Tuan, 2001). Space becomes place over time as a result of sensory experience and related cognition and emotion. For instance, a new arrival in a neighborhood will initially perceive the location as a "blurred image", mere space in which nothing is recognized and thus nothing is defined. Over days, weeks, and months of experience interacting with this neighborhood, however, the images obtain meaning and the neighborhood achieves the status of a perceived "place" (Tuan, 2001).

In at least two cases, IS researchers have found that these concepts have salience in understanding activities in an information technology environment, specifically the virtual world environment (Mennecke et al., 2011; Saunders et al., 2011). We suggest that this theory base can be even more broadly applied to understand users' relationships to websites or other digital content repositories. Thus, a website or other digital content repository becomes a digital place as users experience them and begin to attach meaning to them. For instance, a user having a good experience with a particular blog may result in that user forming positive associations with it and ascribing it meaning. A user having a negative shopping experience on an e-commerce site may also result in attached meaning as that site ceases to be considered formless "space" on the Internet and instead becomes a better-defined "place" (that the user intends to avoid).

Relph discusses that place can exist in both the formal and perceptual (1976). The formal sense of place refers to agreed-upon, defined locations with clear boundaries on which there is general agreement: the Pacific Rim, India, the Elbe River, the Flatiron District, or the Santa Ana Freeway. Each of these names references a location with a generally accepted definition and, thus, each is a "formal place". The idiosyncratic, perceptual sense of place, on the other hand, is dependent on individuals' experiences with locations and the specific meanings that perception and experience attach to those locations: your home, the place where you met your spouse, or the restaurant where you got food poisoning.

We propose that users' interactions online can be understood as having a similar dichotomy. Locations (i.e., sets of content such as a given website) can be understood both as formal or perceptual places. Formal places coincide with domain names, rights holders, and general understandings. For instance, mentioning Amazon.com will generally result in a common understanding of an e-commerce website hosted on that domain. On the other hand, each individual interacts with online locations idiosyncratically and thereby develops an idiosyncratic perception through experience. These idiosyncratic perceptions may or may not track with formal places online.

To further understand how individuals come to attach meaning to locations and, thus, how the sense of place develops, Gustafson built on Relph's conceptualization of experiential place (1976), by further theorizing and testing the existence of sub-dimensions of place, notably self, others, and environment (2001a, 2001b). Here, *self* refers to the personal connections one feels to a place. Such connections may be due to the importance of various life stages experienced, emotional associations, or activities conducted at the location. This may also refer to the connection with a place stemming from self-identification, e.g., a person may derive an identity based on the neighborhood where that person grew up. *Others* refers to the meaning attached to a place as a result of the people who are associated with it, whether the individual has a relationship with those individuals or not. Within the digital platform context, this might help describe the feelings of a gamer toward the "place" of a particular online game where other players behave belligerently; a social networking site like Facebook may be viewed as a place of meaning based on the types of people who frequent it. *Environment*, then, captures the sensory nature of a place, the way it looks, smells, sounds, etc., as well as its location, i.e., proximity to other places.

We suggest that, in the online environment, users come to understand websites and other digital content collections as places in a similar manner. That is to say, that users come to attach meaning to content collections, whether perceived along formal lines or merely perceptual, based on these same three sub-dimensions. Thus, as users have experiences with websites that relate to the self (e.g., a website that helped resolve a marital problem), to others (e.g., the website where a new friendship was struck), or to the environment (e.g., the website with the flashing colors and intense imagery), those websites take on meaning and become online places.

4.2.2. Borders, Digital Locations, and Border Strength

Implicit in the existence of online locations is the concept of *digital borders*, or the specifications around an apportioned quantity of information in a digital environment.

We assert that digital locations, as with physical locations, are specified as a result of these borders. Within the field of IS, this concept dates back at least as far as Kent (W. Kent, 1978), who discussed issues with categorizations of data (i.e., drawing borders around it), in particular with regard to how they effect the ways that data are represented in databases. He warned that, within such data environments, "the boundaries and extent of 'one thing' can be very arbitrarily established" (p. 5) and that "there is no natural set of categories" (p. 13). Similarly, we suggest that there is not necessarily one single way to perceive borders around digital content (e.g., websites) and, thus, that different users may perceive the content with which they interact differently. Some may see and recognize the formal borders, while others may, for various reasons, not recognize them.

However, we also assert that formal borders are related to perceptual borders and, specifically, that the stronger the formal border erected, the more likely it is to be recognized. In the physical world, borders can be made stronger or weaker. Tuan, for instance, discusses political borders and states that "politics creates a place by making it visible" (1975, p. 163), that is, by clearly delineating borders, erecting signs, creating an identity, etc. In other words, formal organizations (e.g., website owners) that want to have locations perceived as places can do so through intentional means, for instance by erecting symbols, providing contrasts, requiring check-points, or through education.

To describe the extent to which formal organizations go about doing this online, we introduce the concept of *border strength*, or the degree to which visual cues, input requirements, sound, etc. mark the transition into a new website. The borders between two neighborhoods in the same city are often relatively weak, with perhaps no clear sign that a transition has taken place. The border between two countries, however, is generally stronger as it often entails a physical barrier (e.g., a gate) and the requirement to stop to show a passport, answer questions from border guards, and make a customs declaration. Similarly, the borders erected by websites can be stronger or weaker and can be a function of design. Some current sites have attempted to strengthen their borders through requiring users to transition onto their sites via an interstitial page that "welcomes" them to the site before facilitating user access to expected content (e.g., Forbes.com, InformationWeek), while others have instituted displays of how many articles the user has accessed in a given month (e.g., New York Times). Other sites try to strengthen borders through striking visual design, interactive features, or through the inclusion of multimedia elements.

Using Tuan's rationale that increased border visibility influences the perception of locations as places, we thus hypothesize:

> *Hypothesis 2: The higher the border strength, the more likely the user will recognize a border.*

4.2.3. Task Structure

The task that drives the user to commence a website visit has been found to be an important predictor of the quality of the website visit. In particular, *task structure*, or "the degree to which the necessary inputs, operations on those inputs, and outputs are known and recognizable to the decision maker" (Browne et al., 2007, p. 91), is a construct that has demonstrated usefulness in a number of previous studies. For instance, it has been found to affect product knowledge increase (Zhenhui Jiang & Benbasat, 2007), perceived website utility (Albert et al., 2004), attitude toward the site (A. Y. Lee et al., 2010), degree of flow experienced (Novak et al., 2003, 2000), and stopping rule usage (Browne et al., 2007).

For this study, we conceptualize task structure as existing on a continuum from low task structure to high. A task with high structure is likely to have a single correct answer that does not require significant abstraction of thought, while a task with low structure is likely to be ambiguous, have multiple possible correct answers, and require the individual to make determinations as to how to evaluate potential answers for appropriateness (D. J. Campbell, 1988).

We expect that in a more closed-ended task (i.e., a task with higher task structure), site visitors will have a more focused approach on their information search and, thus, be concentrating on finding the information point needed rather than on the sites that host the information points. As such, we expect there to be a negative relationship between task structure and the user's ability to recognize a site.

> *Hypothesis 4: The higher the task structure, the less likely the user will recognize a border.*

4.2.4. Attribution and Memory

Finally, we address the concept of attribution. We define *attribution* as the act of assigning credit (or blame) to a given party based on the perceived value contributed to an experience. In the website context, this would reflect a user's crediting a given website for providing a needed solution or sought-after information.

In an environment where deep linking is common and fewer users enter by the "front door" of a website's home page, attribution becomes an important goal for website owners. As has been shown in past research on users' willingness to become (or intend to become) a user of a site, future use depends on perceptions of a site's usefulness (Gefen, Karahanna, & Straub, 2003a; e.g., Gefen et al., 2003b; Lederer, Maupin, Sena, & Zhuang, 2000; Pavlou & Fygenson, 2006; Pavlou, 2003). From an experiential perspective, the perception of usefulness may be predicated on the perception of past experiences with a website or location. Thus, if a user attributes credit to a website for having helped resolve a problem or answer a question, then the user should perceive that site as more useful and thus prefer it.

To illustrate the idea of attribution in the online context, consider a user who is in the market to buy a car. The user prefers one that is economical, but nevertheless fun to drive. The user therefore goes to a search engine and enters "economical car fun to drive", resulting in a search results page of ten links leading to pages with relevant content. The user clicks on the first four of these links, reads the article hosted at each, and, as a result of those experiences on those websites, determines which car best fits the selected parameters of economy and fun. To what degree and to which sites does the user attribute the success of the completed task? One possibility is that the user remembers the experiences exactly as they happened and without bias and, as a result, assigns each site credit based on its actual contribution. In this case, if each of the four sites contributed the same actual amount, then, assuming the user does not credit the search engine itself for a contribution, each site receives 25% of the total credit.

However, attribution is a function of memory, which is subject to distortion and inaccuracy (Belli, Lindsay, Gales, & McCarthy, 1994; Schacter & Dodson, 2001; Schacter, 1999). Memory has long been a subject of interest for psychology researchers. There exist two major perspectives on memory: the quantity-oriented approach and the accuracy-oriented approach (Koriat, Goldsmith, & Pansky, 2000). Under the older, more-established quantity-oriented approach, memory acts like a bookshelf on which memory items can be stored and retrieved (Ebbinghaus, 1913; Roediger, 1980). In this approach, memory is seen as a list of equivalent items and is assessed in terms of the number of items recalled (Koriat et al., 2000). In the accuracy-oriented approach, on the other hand, memory is seen as a reconstruction of experience and is measured in terms of how well the recollection fits actual events (Koriat et al., 2000).

Schacter specifies six ways in which memory can be influenced[11] (1999). These include transience (memories weaken over time), absent-mindedness (lack of attention to detail prevents a memory from being recorded), blocking (memory of the event exists, but cannot be retrieved), misattribution (a memory is attributed to an incorrect time, place, or source), suggestibility (suggestions made during recall influence memory), and bias (memory is distorted by preexisting knowledge and/or beliefs). He further sub-divides misattribution into three sub-dimensions: (1) a correctly remembered fact attributed to the incorrect source (consider those falsely imprisoned based on "eyewitness testimony"), (2) attribution to the self due to the assumption that

[11] In the article cited, Schacter describes seven (as opposed to six) problem items related to memory; however, the seventh, persistence, reflects problems attendant to the inability to forget, particularly in the case of traumatic memories, rather than any defect in accuracy.

the thought was original, rather than one that was experienced (e.g., unintentional plagiarism), and (3) recall of events that did not actually happen.

From among these factors, a number emerge that could cause users who had the same experience to attribute credit for that experience differently. Transience could cause our hypothetical user to forget parts of the experience over time and thus introduce variance in attribution depending on when the attribution was requested. A lack of attention to detail (absent-mindedness) could cause the user to miss the identities of one or more of the sites, thus making attribution impossible. Blocking could occur, for instance if the request for attribution causes the user to panic and be momentarily unable to remember the identities of the sites. All three of Schacter's components of misattribution could also play roles; the user could remember Site A's content having resided primarily on Site B, for instance (sub-type 1), could assume that most of the knowledge gained was something the user already knew (sub-type 2), or could recall having visited sites that were not among the four visited, thus attributing credit to a site that was not part of the experience (sub-type 3). Suggestibility could potentially play a role (if, for instance, the attribution question were to be asked while the user had one of the four sites currently on screen). Finally, the user could be biased in attributing credit as a result of having used Site C with much success in the past or having heard from a knowledgeable friend that Site D is a very authoritative source of car-related information.

Border strength could be reasonably expected to impact two of Schacter's factors: absent-mindedness and misattribution. A strong border reinforces the identity of the website being visited and slows down the user's experience by delaying (e.g., website only loads after waiting 10 seconds), diverting attention (e.g., design cues), and/or requiring additional action (e.g., requiring a click). Further, it ameliorates all three sub-components of misattribution; reinforcing the identity of the site to be visited should lessen the likelihood of the experience being attributed to (1) the incorrect (but experienced) source, (2) the individual, and (3) a source not included within the experience. We expect therefore that higher borders will be related with users understanding where they are and, therefore, attributing the quality of their experience to the site in question.

> *Hypothesis 3: The higher the border strength, the more the user will attribute the quality of an experience to the site.*

We expect task structure will also have a significant relationship with attribution, for reasons that are the inverse of border strength. We expect that users who are performing a high-structure task will be focused primarily on the task. In such a situation, users search for specific elements to complete the task, whereas in a low-structure task users are prone to seek a given volume of information (Browne et al., 2007). This, we expect, will portend to higher absent-mindedness, as users focus their attention on scanning the information at hand for singular, specific data points rather than adopting a more experiential mindset. Given the limitations of memory and the demonstrated negative effect that an increase of data points has on the ability to correctly recall details (e.g., Miller, 1956), this search for more and more concrete data points should make correct recollection more difficult. We therefore expect that a higher structure will result in users being more likely to attribute credit to the wrong site, to themselves, or to websites that may have been outside the scope of their experience. Thus:

> *Hypothesis 4: The higher the task structure, the less the user will attribute the quality of an experience to a site.*

We also expect to see a significant relationship between the ability to recognize a location and the attribution of credit for an experience. The same issues that affect memory in terms of attribution should also affect it in terms of recognition. Thus, sites that are recognized should be more likely to be attributed credit for an experience.

> *Hypothesis 5: Attribution of credit for an online experience will be higher when a website is recognized as having been part of that experience.*

As mentioned, Schacter also includes other possible influences on memory that we do not expect to be affected by border strength, task structure, or border recognition. We thus note the importance of controlling for these factors where possible. While blocking would be difficult to address, transience, suggestibility, and bias can be controlled for to some extent. We will thus ensure that attribution is rendered within a reasonable, known amount of time across users to mitigate the potential effects of transience. Suggestibility issues will be minimized by using random ordering and recording the identity of any websites users reference while assigning attribution.

4.2.5. Hypothesis Model

Taking each of these hypotheses into account yields the following model.

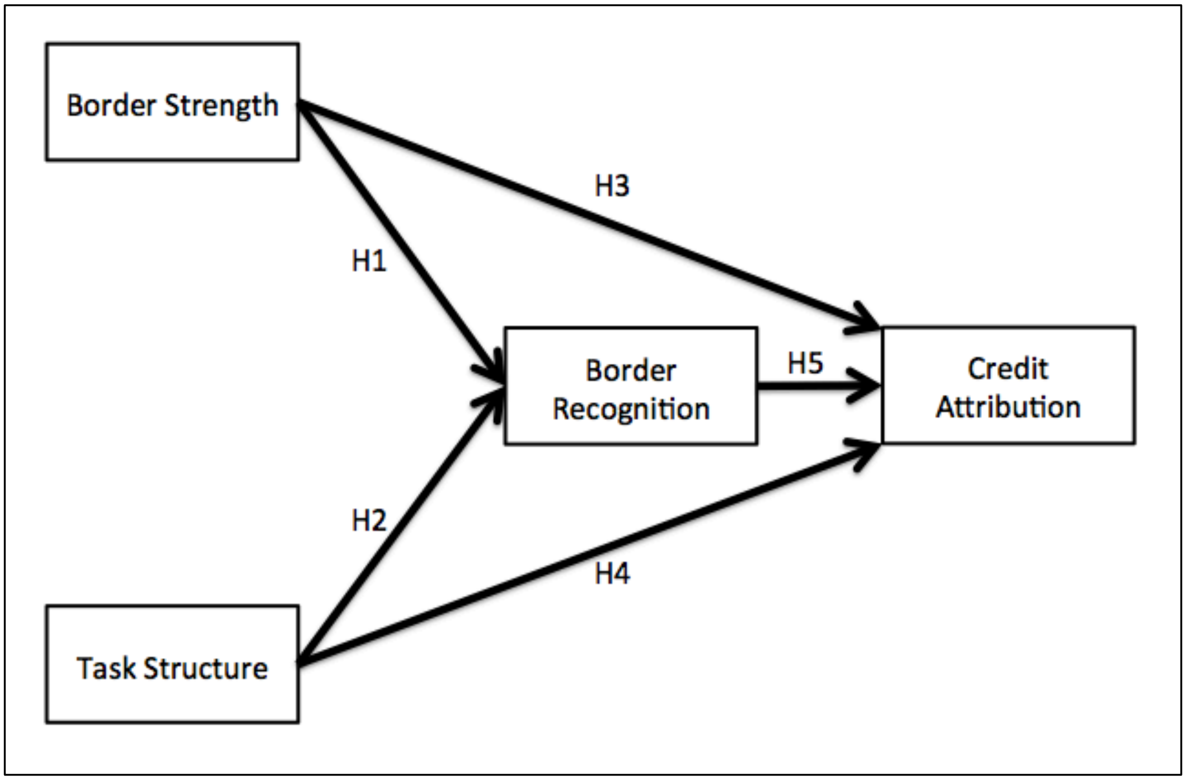


Figure 4.1. Consolidated hypothesis model.

4.2.6. Familiarity, Susceptibility to Interpersonal Influence

To better isolate the strength of the relationships among our variables of interest, we also identify key control variables to be used in our model. First, we control for familiarity. *Familiarity* refers to the amount of experience, direct or indirect, with an entity (Alba & Hutchinson, 1987; M. C. Campbell & Keller, 2003) and reflects associations with an individual's memory (M. C. Campbell & Keller, 2003; Lowry et al., 2008). Though not a variable of interest in this study, we expect that those who have stronger experiences and stronger associations in memory with a website will be more likely to be able to recognize its existence and also more likely to attribute credit to it.

Using the same rationale, we similarly control for participants' direct past experience with the websites.

A user who views a brand as exceptionally important may be more likely to prefer, and thus recognize and attribute credit to, a website hosted by a stronger brand than that hosted by an unknown brand. To this end, we employed the susceptibility to interpersonal influence construct, which refers degree to which individuals feel a social need to affiliate with a given brand through purchase or use (Bearden, Netemeyer, & Teel, 1989). This construct consists of two sub-dimensions, the normative, which reflects "the tendency to conform to the expectations of others" (p. 474) and the informational, which refers to "the tendency to accept information from others as evidence about reality" (p. 474). Thus, we adopted the susceptibility to interpersonal influence construct as an additional control variable.

4.3. Data

4.3.1. Experimental Design

Given our hypotheses, we opted for a 2×3 experimental design that included two levels of structure, low and high, consistent with past IS research (e.g., Browne et al., 2007), and three levels of border strength, low, medium, and high. Both of these variables were operationalized using experimental manipulation. Task structure was operationalized by asking participants to complete one of two news-related online tasks on a purpose-built set of Internet sites. In the high-structure condition, the task required participants to find answers to ten straightforward questions, each of which had only one correct answer. In the low-structure condition, participants were asked to answer five open-ended questions. Border strength was manipulated by the use or non-use of interstitial pages; in this context, an interstitial page is a web page that comes between a user's clicking on a link and the appearance of the actual content related to that link. Our manipulations included a site identity-confirming interstitial page in the high border condition, an interstitial page with a third-party advertisement in the medium border condition, and direct access to requested pages (i.e., no interstitial page) in the low border condition.

4.3.2. Data Collection

4.3.2.1. Sample

Amazon's Mechanical Turk (MTurk) service was used to recruit participants. While this represents a relatively new method of recruiting, earlier research has shown that such subjects yield results that are at least as reliable as those recruited through more typical means, such as undergraduate student recruitment or recruiting services (Buhrmester & Kwang, 2011; Mason & Suri, 2012; Steelman, Hammer, & Limayem, 2014).

A total of 258 participants were recruited. To control for potential nationality-based cultural effects, we limited participants to those living in the United States and US territories. We further specified that participants must have successfully completed 500 MTurk tasks (i.e., they must have had their work approved in at least 500 cases) and have a 95% approval rate (i.e., had their work rejected no more than 5% of the time). Data were collected over a span of 16 days.

From the 258 original participants, ten failed to complete the entire survey and supply the required log file. Of the 248 remaining, an additional seven did not interact with the purpose-built task websites provided and, thus, did not provide usable data, leaving 241 usable user records. Among these, 52.7% were male, the mean age was 34.2, and the median age 32 (see Appendix C, Figure C.1 for age range distribution). Further, 46.4% had bachelors degrees (see Appendix C, Figure C.2 for distribution) and 40 of 50 US states and Puerto Rico were represented[12].

4.3.2.2. Experimental Setting

Because our research questions are particularly relevant to content-based commercial websites and because we needed a context that would cause participants to cross website borders naturally, we decided to use news gathering as our task context. As the effects of subject matter interest level were outside the scope of this study, we further opted to control for this by identifying a subject matter with low variance of interest as determined in an earlier pilot study. In this pilot study, a convenient sample of 21 respondents was asked to rate their level of interest for each of 15 candidate news stories that varied in subject matter from geopolitics to business news to entertainment on a scale from one to seven. Among these 15 candidates, the news story with the lowest standard deviation was (early 2014's) political unrest in Egypt (s.d. = 1.22). However, as this seemed a potentially emotionally charged subject matter that might therefore influence results, we opted instead for the option with the second-lowest standard deviation, namely the (then-) recent Norwegian general elections of 2013 (s.d.

[12] The user sample did not include any participants from the District of Columbia or any other US territories or holdings (e.g., Guam, American Samoa), all of which were eligible to complete the task.

= 1.39). For descriptive statistics of these interest values from our pilot study, see Appendix C, Table C.1.

Given the specific nature of this setting, to enable us to better generalize our results, we controlled for the degree of experience participants reported in using the Internet for news gathering. Those who are more accustomed to using such sites and conducting such tasks may be better (or differently) attuned to the presence of borders and more savvy in their evaluations of such experiences.

With this in mind, we also controlled for task subject knowledge. As mentioned, our data collection required experimental participants to search news stories related to a specific event, namely the 2013 Norwegian general election. We expected that those with significant prior knowledge of Norway and its political events may have a different experience in finding related information, which might impact those users' recognition of borders and attribution of credit for their experiences.

4.3.2.3. Task Stages

The experiment was conducted in three stages presented in successive order: the Pre-Task Stage, the Experimental Task, and the Post-Task Stage. Participants were required through the MTurk system to complete all three stages within two hours. The mean completion time (for all three stages) for the 241 participants was 33 minutes, 18 seconds and the median completion time was 30 minutes, 41 seconds.

4.3.2.4. Pre-Task Stage

In the pre-task stage, participants were asked to complete survey instruments for three constructs: Product Involvement (Zaichkowsky, 1994)[13], Internet News Use, and Task Subject Knowledge (i.e., existing knowledge of the subject matters referenced in the task). All three instruments demonstrated Cronbach's alpha scores over 0.7 and are thus considered to be reliable (Nunnally & Bernstein, 1994), Product Involvement α = 0.948, Internet News Use α = 0.857, Task Subject Knowledge α = 0.727. See Appendix C, Table C.2 for items from the Internet News Use and Task Subject Knowledge instruments, Table C.3 for correlation matrices for these items, and Table C.4 for descriptive statistics.

4.3.2.5. Experimental Task

Following the Pre-Task Stage, participants were asked to complete the Experimental Task. This required them to interact with a purpose-built set of sites to use in answering questions that reflected either a high-structure or low-structure condition. These sites consisted of an invented, static, Google-branded news search results page (see Appendix C, Figure C.3) and nine branded sites, each of which hosted a single article. The search results page included links to unique articles included within each of the nine branded sites. The order of these links was randomly generated for each participant and the ordering was recorded so that our results could be controlled for an article's positioning on this results page (i.e., site display rank).

The identity of the nine branded sites was determined based on a Google search for the term "norway elections"; the nine sites included BBC.com, DW, The Economist,

[13] For this construct, the same instrument was used for this study as in Study 1. See Appendix A, Table A.1 for items included within this instrument.

The Independent, Inter Press Service (IPS), LATimes.com, NBC News, the Nordic Page, and Reuters. The purpose-built branded sites included the actual sites' mastheads as well as a branded footer (see Appendix C, Figure C.4). Each page included a link to "Stories from Other News Sites", which led to the related articles appearing on the other branded sites included within the experiment. The order of these links matched the display order on the search results page.

Nine different news articles related to the 2003 Norwegian general election were either substantially adapted from existing articles (in six cases) or were entirely original (three cases). The nine articles varied in length from 505 to 767 words; readability ranged from 11.0 to 12.5 on the Fleisch-Kincaid grade level readability scale (see Appendix C, Table C.5 for article headlines and statistics regarding the content articles). For each participant, the nine articles were randomly assigned to one of the nine branded sites, thus addressing the potential for sample bias toward a given site based on its hosting of a specific article. For any given site, the same article content was displayed within that site for the entirety of a given user's task. Apparent links from these pages to pages not included within the experiment (e.g., link to the home page) were inactive.

Those selected into the high structure task were asked to answer ten highly specific questions about the election. Participants were instructed to restrict their search to the search results page and the articles to which that page linked. The answers to five of the ten questions (C1-C5) were common to all nine content articles (i.e., any one of the articles could be used to answer these five questions). In order to ensure that participants would need to access multiple sites to complete the task, answers to the other five, “uncommon” questions (UC1-UC5) were each unique to one of the nine articles (i.e., the answer could only be found by reading the one article in which it was contained). See Appendix C, Table C.6 for the text of the ten questions. The mean of correct answers for all questions for those in the structured task condition was 7.98; the mean of correct answers for the uncommon questions (i.e., those found in only one article each) was 3.55. Note that headlines and article synopses (teasers) provided on the search results page were written such that high structure answers could not be found without accessing the article content. Participants in the high structure task group took a mean of 16 minutes, 24 seconds to complete the task.

Those assigned to the low structure task group were asked use a provided link to the purpose-built search results page and the articles found there to research the 2013 Norwegian general elections for ten minutes and then answer five questions. These five questions (see Appendix C, Table C.7) represent a low-structure task (Browne et al., 2007) in that they have multiple possible correct answers and require the answerer to develop a framework for answering. The mean number of words used in answering the five questions was 132.4 (median = 116). Participants in the low structure task group required a mean of 15 minutes 31 seconds to complete the task.

In both the high and medium borders conditions, participants were shown an interstitial page prior to being provided access to article content. Thus, when the participant clicked on a link on the search results page (or on an article page), that participant would be taken to an intermediate page and instructed to either click to continue to the content page or wait 10 seconds for the content page to appear. This approach reflects actual practice on some content websites (e.g., InformationWeek, Forbes). In the high border condition, the content of the interstitial page reinforced the identity of the site that was loading; in other words, if the participant clicked on the BBC News link, the interstitial page included a large BBC News logo and the "Continue" link included the name of the site. In the medium border condition, the participant was shown a third-party advertisement[14] and no reference was made to the site being loaded. Please see Appendix C, Figure C.5 for samples of both the high- and low-border interstitial pages.

To capture participant activity during the task, participants were asked to record their activities using the Problem Steps Recorder (PSR) software found within the Windows 7 and Windows 8 operating systems. Once the task was completed, participants were asked to then upload the file generated by the PSR to the survey software environment used in the experiment. In addition, the database used to dynamically display data to participants also kept a record of which articles were shown on which branded sites and in which order these articles were presented to users on the search results page.

4.3.2.6. Post-Task Stage

After completing the task and uploading the PSR log, participants were asked to complete additional survey items from which variables were derived for site

---

[14] These third-party advertisements were randomly selected from a pool of 15 different 640x480 pixel advertisements for real brands that were specially adapted for this purpose. The advertisements did not link out to the brands' websites.

recognition (i.e., recognizing having visited a site), credit attribution, cognitive absorption, and susceptibility to interpersonal influence (SII). For recognition, subjects were given the name and shown an image of each of the nine branded sites and asked whether they had used that site as part of the task. They were then asked to attribute value to each site in terms of its helpfulness in completing the task; a guideline was given (though unenforced) that the attributed values for all sites taken together should equal 100.

The cognitive absorption instrument used was nearly identical to that in Agarwal and Karahanna (2000), in which the construct is considered to exist in five sub-dimensions of temporal disassociation, immersion, enjoyment, control, and curiosity. To mitigate survey fatigue, however, we used only the three highest loading items (as reported by Agarwal and Karahanna) for each sub-dimension. All sub-dimensions had sufficient reliability based on Cronbach's α (Nunnally & Bernstein, 1994): temporal disassociation sub-dimension $\alpha = 0.878$, focused immersion sub-dimension $\alpha = 0.748$, enjoyment sub-dimension $\alpha = 0.897$, control sub-dimension $\alpha = 0.761$, and curiosity sub-dimension $\alpha = 0.854$.

The susceptibility to interpersonal influence instrument was developed by Bearden et al. (1989) and was used here verbatim. This instrument measures two sub-dimensions, the normative (eight items) and the informational (four items). Here again, both sub-dimensions demonstrated sufficient reliability using Cronbach's α: normative $\alpha = 0.931$ and informational $\alpha = 0.813$[15].

In addition to these, participants were asked to assess their familiarity with each of the nine sites using three seven-point Likert scale items. This instrument also demonstrated sufficient reliability, Cronbach's $\alpha = 0.910$[16]. Finally, participants were asked to report how frequently they had used each of the sites prior to the experimental task. For descriptive statistics by site for both familiarity and frequency of use, see Appendix C, Table C.10.

4.3.3. Data Coding

Data were then assembled and coded. As a result of the data collection procedure, data existed in three different locations: (1) the survey results in the Qualtrics environment, (2) data stored in the database located on the website server, and (3) the participant PSR logs. The survey data were straightforward and could be easily downloaded in comma-delimited format, with each participant's survey responses clearly available without significant additional handling.

Website database data included information on which articles were associated with which websites for each user as well as the order in which the site-article pairings were displayed to each user. These data were keyed on a session identification number that the participants had been asked to retrieve as part of the survey; this number was used to match survey data to the server database data.

---

[15] Factor loadings for both Cognitive Absorption and Susceptibility to Interpersonal Influence instruments are provided in Appendix C in Tables C.8 and C.9 respectively.
[16] For a correlation matrix of familiarity and frequency measures, see Appendix C, Table C.10.

The PSR log data required more handling to extract and code. PSR logs are XML formatted (see Appendix C, Figure C.6 for a sample from this study). These files were transformed into tab-delimited files using a Perl script written by the researchers expressly for this purpose. At this level, each record of data consists of a sequence number, a time stamp, a user action (e.g., a click, text input, mouse scroll wheel use), and a window title (i.e., the words included at the top of a website window). See Appendix C, Figure C.7 for a sample of this action-level file. These data were then parsed further to create user-level records in which were recorded the number of interactions (e.g., a page load, a mouse wheel use) for each user-site dyad (nine dyads per user).

The data from these three sources were then recombined into a single spreadsheet using another Perl script. This final spreadsheet included data at the user-site dyad level, allowing us to proceed with the analysis using a multi-level mixed regression model.

4.4. Analysis

To test our hypotheses, we specify two multilevel models, one for each of our dependent variables, border recognition and credit attribution. Since our unit of analysis was the user-site dyad, multilevel models were appropriate as the parameters of interest varied at both the user and site levels.

In the first model, we predicted whether the participant recognized having used a site and, thus, our dependent variable was binary and required a multilevel mixed effects logistic regression approach. Only dyads in which the participant interacted with the site were used in the regression and, thus, n = 1,461, a mean of 6.06 sites per user. The results of this regression can be found in Table 4.1 below. Means and descriptive statistics for the individual-level variables included in this model can be found in Appendix C, Tables C.12 and C.13 respectively.

Variable	Coeff.	S.E.	z	p	Result
Border Strength	0.433	0.147	2.95	0.003**	H1 Supported
Task Structure	0.770	0.249	3.09	0.002**	H2 Not Supported
Site Interactions	0.049	0.016	3.01	0.003**	
Task Time	0.0002	0.000	1.13	0.258	
Cog. Abs.[17]	0.216	0.092	2.34	0.020*	
S.I.I.	-0.257	0.093	-2.76	0.006**	
News Use	0.063	0.157	0.40	0.687	
Task Subj. Knowledge	-0.028	0.122	-0.22	0.822	
Site Display Rank	0.010	0.030	0.33	0.740	
Site Familiarity	0.168	0.100	1.67	0.095	
Frequency of Site Use	0.172	0.111	1.56	0.120	
Constant	-1.975	1.081	-1.83	0.068	

Table 4.1. Results of mixed effects logistic regression estimation for site recognition. Random effects estimate for the user was 1.34, S.E. 0.141. Mixed effects logistic regression: Wald χ^2 = 78.02, p < 0.001. LR test of marginal logistic regression: χ^2 = 109.32, p < 0.001; * = significant at α = 0.05, ** = significant at α = 0.01.

The data support Hypothesis 1, that higher border strength results in a higher probability of users recognizing their locations, and, thus, the borders between locations. Surprisingly, the relationship between task structure is not only in the opposite direction of our expectations, but is also significant: within our data, the higher the task structure, the more likely participants were to recognize having visited sites. This may have been the result of high-structure participants frequently checking and re-checking the same sites in order to find the specific information needed to answer questions, whereas those in the low structure condition could read an article one time to

[17] In the actual estimation, only the curiosity sub-dimension of cognitive absorption was found to be significant in predicting either border recognition or attribution; thus, the cognitive absorption variable noted in Tables 4.1 and 4.2 consists only of the curiosity sub-dimension. Similarly, only the normative sub-dimension of the susceptibility to interpersonal influence was found to be significant in these two estimations and, thus, the S.I.I. variable named in these tables reflects only that sub-dimension.

form an opinion, and did not need to necessarily re-check the site thereafter. Further analysis showed no significant evidence of an interaction effect.

We then estimated our mixed-model regression addressing user credit attribution. Similar to the first regression, this regression only included those user-site dyads where credit attribution was given, thus n = 1,068, a mean of 4.43 sites per user. The Likelihood Ratio (LR) test of the mixed-effects linear regression did not reject the null hypotheses that there were no significant random effects at the user level. Therefore, for predicting credit attribution we utilized an ordinary least squares regression specification. We derived robust standard errors to test our hypotheses by accounting for the heteroscedasticity of credit attribution by users. The results of this estimation are given in Table 4.2.

Variable	Coeff.	S.E.	z	*p*	Result
Border Strength	0.014	0.006	2.23	0.026**	H3 Supported
Task Structure	-0.062	0.011	-5.43	<0.001**	H4 Supported
Border Recognition	0.085	0.013	6.73	<0.001**	H5 Supported
Site Interactions	0.003	0.001	3.08	0.002**	
Task Time	-0.00002	< 0.001	-2.08	0.037*	
Cog. Abs.	0.001	0.004	0.21	0.833	
S.I.I.	0.005	0.004	1.29	0.198	
News Use	0.005	0.007	-0.69	0.491	
Task Subj. Knowledge	-0.005	0.005	-0.94	0.349	
Site Display Rank	-0.004	0.002	-1.91	0.056	
Site Familiarity	0.010	0.006	1.51	0.132	
Frequency of Site Use	0.018	0.007	2.65	0.008**	
Constant	0.020	0.051	0.38	0.685	

Table 4.2. Results of mixed regression estimation for site credit attribution. $F(12, 1055) = 19.01$; Prob > F <0.001; Adjusted R-Squared=0.17; * = significant at α = 0.05, ** = significant at α = 0.01.

Here, we see significant support for all three hypotheses tested. The higher the border structure, the greater the attribution of credit to the site (H3). Conversely, users completing a higher structure task were less likely to attribute credit to a site than those completing a lower structure task (H4). Finally, a participant's recognition of having used a site significantly predicted that participant's attribution of credit to that site for the task completed (H5).

We tested if border recognition mediated the effects of border strength and task structure on credit attribution using the Baron and Kenny analysis (1986) and using the Sobel test (1982). The mediation analysis revealed that border recognition was a statistically significant partial mediator variable between border strength and credit attribution (Sobel statistic=2.4; $p < 0.05$), and between task structure and credit attribution (Sobel statistic=2.6; $p < 0.01$). This statistically significant partial mediation by border recognition indicates the importance of the causal relationships between border strength (and task structure) and the ability of users to recognize a website, and eventually their credit attribution to a website.

4.5. Discussion

4.5.1. Implications for Practice and Research

Our models yield interesting results. With regard to the importance of website border strength, support for H1 and H3 indicates that users are significantly more likely both to recognize sites used and to attribute credit to sites for tasks completed when those sites have higher borders strength. For practitioners, this indicates borders can play a positive role both in increasing the likelihood of users remembering the site as well as in users attributing successful task completion to the site. With this in mind, site owners should investigate using design elements to erect higher, clearer borders to more obviously demarcate a user's point of entry. This could be achieved through interstitial pages as was done in this experiment, or through other design-based means such as animations, layouts, welcome chimes, etc. Additional research in this area would be helpful to determine the most effective form of border.

We also caution, however, that there may be downsides associated with erecting higher borders; we suggest that further investigation into potential negative effects of increased border strength measures (e.g., annoyance, distrust) is necessary to fully understand the dynamics involved with website borders and better illuminate the parameters around the related practitioner decision-making.

With regard to the effects of user task structure, our data reveal mixed findings. H2 was not only not supported, but a significant result was found in the opposite direction: participants entering a site with the high-structure task yielded higher levels of site recognition than those with the low-structure task. Given the nature of the high-structure task, and that half of the questions could only be found in one each of the nine articles presented, however, this result may be an artifact of our experiment design, which encouraged those in the high structure group to visit the same site repeatedly in some cases. However, there may also be reason to believe that this result is an important one. For instance, those individuals who come to a site intending to complete a more-structured task may focus more strongly on the identity of sites used to evaluate their usefulness for future needs. Given that this result runs counter to established theory, we believe that further research should be conducted to validate this finding.

In terms of attribution, however, we note that task structure again plays a significant role. Those who use a site to complete a more-structured task are less likely to attribute credit to that site than those who come to complete a less-structured task. This suggests that those looking for specific information to answer specific questions may be more prone to frustration with a site — an idea further supported by the fact that higher structure yielded higher recognition, indicating that in many cases sites were recognized, but not given credit for helping with task completion. With this in mind, websites may do well to better facilitate information finding for high structure task site visits in order to minimize any potential negative impression that could be made on these visitors.

We also note another interesting result implicit in findings: in fact, some users *do not* recognize the locations they visit and the borders that form them. Further, this study used a measure of recognition to indicate user memory; this recognition is measured by presenting participants with an entity (in this case the website) and asking if they remember having used it. This represents a lower threshold of memory than would have "recall", in which participants would be required to recall the identity of sites from memory and without prompting (e.g., Bagozzi & Silk, 1983). Had recall been used instead, user forgetfulness of location would likely have been even more apparent.

Attribution is another potential point of interest here. Our study did not investigate whether attribution was necessarily given to the site that provided useful information, but, rather, merely asked participants to which sites they attribute credit for completing the task. Looking in particular at those in the high structure task group, where a participant would have had to look at a minimum of six sites to answer all ten questions correctly, it is interesting that fewer than six are credited by most participants.

Clearly, then, credit attribution is being withheld from some sites that were helpful in completing the task. Focusing specifically on this question of when attribution is given to the site that actually helped solve the problem and when it is “misattributed” would be an interesting extension of this study. We note that, in addition, further discussion of the concepts of correct and incorrect attribution would be a potentially profitable enterprise for future researchers.

Establishing this lack of recognition and attribution has potential importance for research. As mentioned, much of the research undertaken in the area of website interactions has used an important assumption to facilitate testing of hypotheses, namely that websites are discrete entities understood as such. These findings suggest that this may not be the case and indicates that there may be room to build upon such past studies by introducing this new, potentially interesting consideration. What impacts might vague site borders have on such constructs as usability, trust, privacy, purchase intent, or intent to use within the online context?

These findings also suggest the need for additional research of border strength effects in other digital content contexts. One of the central features of the Internet and, indeed, this study is the movement of users from one site to another through online links. In a sense, similar movement happens within other digital platforms. Those using smart phones or other mobile platforms move from app to app, sometimes consuming content from different third-parties within the same app. In the context of video streaming devices, users move from Netflix content to Amazon Prime content to YouTube content. How do users perceive borders in these contexts and what implications does that perception have for both platform owners and content developers?

4.5.2. Limitations

We note that our study is subject to some limitations. For one, because of our reliance on the PSR software available on Windows 7 and Windows 8, our sample was limited to MTurk participants who had access to one of these two operating systems. As such, our sample may have omitted users of Apple's OS X, Linux distributions, and users of non-PC devices. Further, two respondents used touch-screen devices to complete the task; the PSR was ineffective in capturing on-site behavior for touchscreen users and, thus, these two cases were omitted from the data set. Given the omission of these users, it may be useful to replicate this study in a way that would allow their inclusion. That said, we would expect similar findings. For instance, within this data set we were able to test for an effect based on the specific browser used (whether Firefox, Google Chrome, or Internet Explorer) and saw no significant effect.

Next, we acknowledge that, as with most experimental samples, there may have been sample bias caused by our use of the MTurk system. However, we note that our respondents showed substantial variance in terms of age, education, and geography within the US, thus suggesting a more representative sample of U.S. adult Internet users than may have been available through other means (e.g., undergraduate student participants).

Finally, as with most "laboratory" experiments, the task assigned participants in this study was contrived for the purpose of the study and, thus, may not have facilitated participants' behaving in a natural manner. For instance, users may not have as many as ten specific questions to which they need answers when investigating news stories, or may not feel the need to interpret events as requested of users in the low-structure condition. That said, the manipulations of this experiment indicate the presence of phenomena, that there is indeed a case where users' recognition of website borders and locations is imperfect, and attribution may vary based on that recognition as well as border strength and the structure of the task to be completed. Further studies in settings deemed more natural may be useful to confirm these findings.

4.5.3. Conclusion

This study makes a significant contribution to the IS literature. First, we introduce into the literature the conceptualization of the Internet as a form of geography to which the concepts of space and place may be applied (Buttimer, 1976; Relph, 1976; Tuan, 1977). We find that, indeed, users in some cases do not know where they are online and do not necessarily recognize the websites they have visited. This suggests that conclusions about how a website experience influences a user's attitudes and intentions toward a website may be presumptuous without considering the user's interactions with that site within the context of a broader online experience. As a result, website owners need not only to consider the degree to which certain constructs relate to a user's intent to return to a website or intent to purchase, but also must consider whether their site is discretely recognized in the first place. Relatedly, we introduce the concept of border strength and show how varying the strength of borders can have a significant effect on online location recognition.

Further, we novelly applied the concepts of memory, in particular those affecting memory accuracy (e.g., Schacter, 1999), to understand how users may attribute qualities of a multi-site online experience to individual websites. We find again here the importance of border strength, as well as task structure (user intent) in determining how users attribute success to sites visited. Interestingly, we can significantly predict user attribution of credit without also considering where users actually found the sought-after information included in the task. Our findings suggest that the way research approaches user-website and user-digital content interaction may be more fully understood by considering how a specific user-website dyad fits within the context of a broader online experience.

Conclusion

These three studies form an important contribution to the academic literature as well as to practice. Previous literature has considered user behavior on websites by isolating user intent and/or by isolating the user's experience to a single website (or looking at the experience on the entire Internet as a whole). These studies highlight the importance of the user element in determining the success of websites and other online content platforms. Further, it shows the potential that exists when we begin relaxing the two aforementioned constraints and consider the possibilities of different intents as well as a website's position as a stopping point within the broader online experience of a user's multi-destination journey.

First, we showed that variation in the actions of users influence the degree of success for brand-focused websites. While this is an intuitive fact in the case of e-commerce sites, where the user acting to complete a purchase is clearly tantamount to success, we saw that user behavior, measured in terms of time and (novelly) territory foraged, mattered in terms of affecting brand engagement as well. Further, we showed that this behavior can be manipulated through the design qualities of the website, particularly as those design qualities affect the user perception of the site's cognitive absorptive and interactive properties.

Our studies also found that user intent influences the way users interact with websites. By testing both high- and low-task structure intent against the same metrics and same sites, we found that these intents affected both the manner in which users foraged for information as well as their ability to recall websites and attribute credit for the quality of their online experiences. We also found that a user's intent can be inferred upon arrival at a website, showing this was possible for users arriving via sponsored search links.

Finally, we showed that viewing an individual website within the context of a broader online experience matters. This was demonstrated in terms of both a user's ability to recognize having visited a website as well as the user's willingness to attribute credit to it for the quality of an experience. Further, we demonstrated how design decisions, which we described in terms of "border strength", could affect both recognition and attribution and, thus, more clearly demonstrate a website's presence and pertinence within a broader online experience.

Along the way, we also made other important contributions to the literature. We reconciled the seemingly discordant perspectives of information foraging theory and brand exposure by introducing the idea of foraging *territory* in addition to foraging time, showing that both efficient within-site travel *and* greater coverage of on-site territory are concurrently important. We introduced into the IS literature the idea of brand engagement as an important dependent variable for website owners, especially given that nearly all types of websites (e.g., e-commerce, information dissemination, branding) rely to some degree on the strength of their brand.

Then, in our second study, we also contributed a novel typology of sponsored search query terms, taking an almost infinite array of potential keywords and distilling them into manageable queries. We also created a novel method for determining the range of possible user intent by applying the idea of the purchase funnel.

Finally, we introduced the concept of digital borders within digital content environments (specifically websites). We showed that borders can be stronger or weaker and that this can have a profound influence. We also introduced the application of the theory of space and place to website interactions and theorized that there are different ways in which an online "space" can become a known and familiar "place" as users attach meaning through their experiences and other means.

Our relaxing of the two assumptions, the isolation of user intent and viewing of websites as discrete entities, creates a number of opportunities for research to expand its understanding of online and digital content interactions. We expect that established constructs will gain new nuance and depth when viewed with these relaxed assumptions. We also expect that the concept of the online world as a geography will open new paths for exploration. In fact, the research presented here suggests a number of future avenues of inquiry.

For instance, future research may usefully endeavor to identify and test the website design attributes and functions that best address the dual need to maximize foraging territory while minimizing foraging time. Different presentation formats (e.g., bullet points instead of paragraphs, imagery instead of text), for instance, may enable both of these to happen at once. Further, we have seen a growing trend toward touch-screen interactions, both through tablets and PC devices. Optimizing the way websites handle this, as well as other input and display types that may emerge, may also play a role in improving user foraging behavior from the website owner's perspective.

Our exploration into website foraging also suggests a usefulness in delving more deeply into user foraging behavior to identify, typify, and then understand more-nuanced patterns and then investigate their implications. Study 1 used a fairly simple approach to measuring foraging in which all page views were counted equally. However, a deeper investigation into the patterns used in terms of types of pages visited, depth in the site, amount of time spent on each page, etc. may yield interesting insights into the varieties of foraging behaviors exhibited and, potentially, what antecedents may relate to those behaviors.

In addition, having brand awareness as a primary goal for websites implies the need to properly measure user interactions to ascertain the degree to which websites meet that goal. Further, traffic generation schemes, such as the sponsored search activities examined in the second study here, can also be optimized by correct measurement of user behaviors. However, doing so requires a specific study into the proper measurement of website-based user activities based on the brand engagement objective. Our findings regarding foraging imply that the correct answer lies somewhere in the maximization of territory foraged and the increase of foraging efficiency. We expect, however, that other factors and nuances may explain more of the variance observed in users' brand engagement deltas. For instance, future research may determine that not all territory can be valued the same (e.g., navigational pages may be worth less than product detail pages or white papers) or that lingering on certain pages may, in fact, be a positive thing from a brand viewpoint.

We also see fantastic potential for inferring user intent based on means of arrival. In both the first and third studies, we saw the importance of user intent in terms of users' foraging behavior and their abilities to recognize and attribute credit to websites. This implies that the optimal user experience on a website may be different based on the user's intent. As demonstrated in our second study, this intent can be inferred based on the user's means of arrival. Future research may better refine this inference, both in terms of sponsored search specifically as well as other avenues of arrival (e.g., direct URL entry, natural search, third-party links, etc.). With more reliable ways to infer intent, further investigation into designs and technologies to better address that intent will become especially important.

Relatedly, in our exploration of the massive data set in Study 2, we viewed intent on somewhat simplistic terms: users were either at the bottom of the funnel or higher-up. We also found that the majority of search sessions did not culminate in purchase and that, therefore, a likely large majority of query-click dyads were not from users at the bottom of the funnel. This, of course, raises the question of whether we can infer user intent at a deeper granularity, particularly for those in the non-purchase stages. Our findings suggest that the broadest, least-specific terms should relate to the highest levels of the purchase funnel (i.e., product awareness). Further study could confirm or refute this theoretical expectation.

Our finding that users are not necessarily aware of their online locations suggests that there may be triggers that cause this recognition. In our study, it was shown that the use of stronger borders in the form of interstitial pages aided later recognition. However, this raises another question regarding the point at which a user becomes aware of having crossed a border. Does it happen at the time of the interstitial page, or does the presence of the interstitial page slow down the user's process, thus enabling the user to focus attention more closely on the website that appears next? Answering this question could give further insights into what makes a border effective.

This could be particularly useful to know given the potential for border strength to come coupled with a negative impact on the user experience. In our third study, we showed that border strength was positively related to a user attributing credit to a site. This suggests that any annoyance caused by the interstitial pages shown was more than offset by the user's increased association of the quality of an experience with that site. However, it seems intuitive that at some point the strength of the border begins to have negative consequences — for instance if the following page could not be loaded from the interstitial for several minutes or if a site used garish font choices and colorful overlay graphics to reinforce its identity and thus rendered its actual content difficult to read. A further study could determine the point at which returns on border strength become negative. This may happen based on whether the user knows the site before visiting it, the nature and urgency of the task completed, and the nature of the border itself. It may be that some borders cause little annoyance, but are very effective at increasing the user's ability to recognize a site and attribute credit to it.

In addition to these avenues of study, we suggest that our findings may extend beyond the limits of the browser-based World Wide Web. Since the popularization of the Internet in the 1990s, online connectedness has spawned new industries (e.g., social networking, online gaming) and consumed all or parts of others (e.g., digital music, video dissemination). Our findings have potential implications in these non-Web contexts as well in exploring user perceptions of the interaction among other Internet-enabled platforms as perhaps influenced by user intent and the strength of various borders. Not only does this dissertation contribute meaningfully to the Web context, but it also provides a starting point for deliberations on these other digital content contexts as well.

In addition to these contributions to research, our studies also present a number of important implications for practice. Website designers in particular are provided a number of ideas for consideration. In showing the concurrent importance of both visit efficiency and territory covered in a website visit, a significant design challenge is brought forward. Web designers need to find ways to use information scent and perhaps integrate new interaction types and technologies to enable this seemingly conflicting requirement.

Designers will also want to pay attention to our findings regarding borders and border strength. As the Web has continued to mature, fewer and fewer website visitors are entering those sites through the "front door" of the sites' homepages. As such, designers need to take into account methods for strengthening the perceived borders of their sites to ensure that users are aware of their online location and can correctly attribute credit to a site that deserves it.

From a more programmatic standpoint, we showed not only the importance of user intent, but also the possibility for inferring intent based on information related to a user's arrival on the site. Programmers and designers may need to work together to develop ways to use this on-arrival information to deliver a website experience tailored to the needs of the user based on the user's inferred location in the purchase funnel.

Finally, this study bears significant implications in terms of measurement. Website-owning firms need to consider the overall objectives of their sites and measure the behaviors and actions of users appropriately to evaluate success and make improvements. Based on our findings, both efficiency and territory covered need to be taken into account to ensure appropriate measurement. We also find the importance of inferring user intent; measurement of this inferred intent across various traffic generation tactics would be useful in better targeting and servicing future website visitors. Finally, our third study showed that in some cases users are unaware of having interacted with a given site. In this case, the site needs to consider the value of such users, necessitating further study into understanding more the conditions that lead to a user's later inability to recognize or attribute credit to the site.

We feel the contributions of this dissertation are vital and timely and hope these studies are found to positively influence the present and future discourse on digital content interactions. We look forward to not only addressing some of the questions and challenges presented by this work ourselves, but also to seeing how other go about resolving them.

Appendix A

Construct	Abbrev.	Item	Source/Basis	Adaptation(s)
Site Visit Intent	sv1	It is very likely that I will (return to) visit…	Coyle & Thorson (2001): "It is very likely that I will return to this site."	Re-wrote to avoid asking respondents to read entire question three times (once for each brand).
Site Visit Intent	sv2	The next time I need a new car, I will visit...	Coyle & Thorson (2001): "I will return to this site the next time I need a (product)."	Re-wrote to avoid asking respondents to read entire question three times (once for each brand).
Site Visit Intent	sv3	Suppose that a friend called you last night asking for advice in finding a new car. Would you recommend visiting...	Kim & Biocca (1997): "Suppose that a friend called you last night to get your advice in his/her search for a (product). Would you recommend him/her to visit (brand's) web site?"	Changed to remove gender pronouns. Also re-wrote end of question to avoid asking respondents to read the entire question three times (once for each brand).
Brand Purchase Intent	bpi1	It is very likely I will some day buy an automobile from...	Coyle & Thorson (2001): "It is very likely that I will buy (brand)."	Included term "some day" in order to avoid implying that it needs to happen in the very near future. Changed end of question to avoid asking respondents to read the entire question three times (once for each brand).
Brand Purchase Intent	bpi2	The next time I need a new car, I will buy...	Coyle & Thorson (2001): "I will purchase (brand) the next time I need a (product)."	Note: Omitted Coyle & Thorson's item "I will definitely try (brand)" as this seems an odd thing to ask about a car. Could perhaps change "try" to "test-drive" or similar in order to include.
Brand Purchase Intent	bpi3	Suppose a friend called last night for advice on a car to purchase. I would recommend that my friend buy...	Kim & Biocca (1997): "Suppose that a friend called you last night to get your advice in his/her search for a (product). Would you recommend him/her to buy a (product) [illegible]	Changed to remove gender pronouns. Also re-wrote end of question to avoid asking respondents to read the entire question three times (once for each brand).

Brand Familiarity	bf1	For each car manufacturer below, please estimate when you last saw an advertisement for the company or one of its vehicles. ("Never" to "Within the Last Hour")	Stewart (1992), Kent & Allen (1993, 1994) --> Brands that advertise are more familiar.	Developed for this paper.
Brand Familiarity	bf2	I personally know a lot of people who own a car made by...	Campbell & Keller (2003) --> Familiar brands may have been tried, family or friends may have used them and told the consumer about them, may have seen ads, or have read about in the press.	Developed for this paper.
Brand Familiarity	bf3	I have read a lot of articles about the company and/or the cars they make.	Campbell & Keller (2003) --> Familiar brands may have been tried, family or friends may have used them and told the consumer about them, may have seen ads, or have read about in the press.	Developed for this paper.
Brand Familiarity	bf4	I have significant personal experience with cars made by...	Campbell & Keller (2003) --> Familiar brands may have been tried, family or friends may have used them and told the consumer about them, may have seen ads, or have read about in the press.	Developed for this paper.
Brand Familiarity	bf5	I know a lot about the company and the cars they make.	Alba & Hutchinson (1987) --> Brand familiarity reflects level of direct and indirect experience.	Developed for this paper; similar to Lowry et al. (2008) "I am not familiar with this type of service"; in our pilot, this wording scored better than one that asked about familiarity directly.
Brand Familiarity	bf6	I could talk about the company for a long time.	Lowry et al. (2008) --> This is a service that I could talk about for a long time.	Adapted to refer to the company (rather than "the service").
Brand Familiarity	bf7	I understand the company well enough to evaluate it.	Lowry et al. (2008) --> I understand the information well enough to evaluate the brands.	Adapted for the context of evaluating the brand.

Brand Attitude	ba1	Please indicate how you feel about (brand). (7-point bipolar from "dislike" to "like").	Li et al. (2002) used 7-point bipolar scale from "dislike" to "like". Gardner (1985), Machleit & Wilson (1988) used "dislike very much" to "like very much".	No changes.
Brand Attitude	ba2	Please idicate how you feel about (brand). (7-point bipolar from "bad" to "good").	Gardner (1985), Li et al. (2002) used 7-point bipolar scale from "bad" to "good".	No changes.
Brand Attitude	ba3	Please indicate how you feel about (brand). (7-point biploar from "unappealing" to "appealing").	Li et al. (2002) used 7-point bipolar scale from "unappealing to appealing".	No changes.
Brand Attitude	ba4	Please indicate how you feel about (brand). (7-point bipolar from" unpleasant" to "pleasant").	Gardner (1985), Li et al. (2002) used 7-point bipolar scale from "unpleasant" to "pleasant".	No changes.
Brand Attitude	ba5	Please indicate how you feel about (brand). (7-point bipolar from "unattractive" to "attractive").	Li et al. (2002) used 7-point bipolar scale from "unattractive" to "attractive".	No changes.
Brand Attitude	ba6	Please indicate how you feel about (brand). (7-point bipolar from "boring" to "interesting").	Li et al. (2002) used 7-point bipolar scale from "boring" to "interesting".	No changes.
Product Involvement	pi1	To me, cars are… (7-point bipolar from "important" to "unimportant")	Zaichkowsky (1994)	No changes.
Product Involvement	pi2	To me, cars are… (7-point bipolar from "boring" to "interesting")	Zaichkowsky (1994)	No changes.
Product Involvement	pi3	To me, cars are… (7-point bipolar from "relevant" to "irrelevant")	Zaichkowsky (1994)	No changes.
Product Involvement	pi4	To me, cars are… (7-point bipolar from "exciting" to "unexciting")	Zaichkowsky (1994)	No changes.
Product Involvement	pi5	To me, cars are… (7-point bipolar from "mean nothing" to	Zaichkowsky (1994)	No changes.

		"mean a lot")		
Product Involvement	pi6	To me, cars are… (7-point bipolar from "appealing" to "unappealing")	Zaichkowsky (1994)	No changes.
Product Involvement	pi7	To me, cars are… (7-point bipolar from "fascinating" to "mundane")	Zaichkowsky (1994)	No changes.
Product Involvement	pi8	To me, cars are… (7-point bipolar from "worthless" to "valuable")	Zaichkowsky (1994)	No changes.
Product Involvement	pi9	To me, cars are… (7-point bipolar from "involving" to "uninvolving")	Zaichkowsky (1994)	No changes.
Product Involvement	pi10	To me, cars are… (7-point bipolar from "not needed" to "needed")	Zaichkowsky (1994)	No changes.

Table A.1. Study 1 pre-task survey questions.

	bf1	bf2	bf3	bf4	bf5	bf6	bf7
bf1	1.000	0.774	0.473	0.512	0.579	0.411	0.473
bf2		1.000	0.514	0.577	0.568	0.423	0.491
bf3			1.000	0.513	0.650	0.662	0.662
bf4				1.000	0.580	0.515	0.530
bf5					1.000	0.663	0.731
bf6						1.000	0.684
bf7							1.000

Table A.2. Brand Familiarity item correlation matrix.

Brand	Abbrev.	Question
Toyota	Qprice	What is the "starting price" (i.e., list price exclusive of options or upgrades) of the Toyota Tundra Double Cab?
Toyota	Qsize	In liters, what is the size (displacement) of the engine available with the Toyota Camry Hybrid LE?
Toyota	Qdealer	What does Toyota recommend as the closest Toyota car dealership for someone living in zip code 92691?
Mitsubishi	Qwarrant	What is the maximum number of years covered under warranty on the Mitsubishi i-MiEV's main drive lithium ion battery?
Mitsubishi	Qseats	How many people does Mitsubishi suggest its Lancer Sportback can seat?
Mitsubishi	Qmirror	What price does Mitsubishi list for the chrome side mirror covers accessory available for the Outlander Sport?
Seat	Qcolors	How many color options are available on the Seat Ibiza SC FR?
Seat	Qmpg	What is the official estimated gas mileage (mpg) range in "urban" driving for the Seat Altea XL?
Seat	Qcity	In which city did Seat open its first manufacturing plant in 1953?

Table A.3. Structured condition questions.

	n	Range	Min.	Max.	Mean	Std. Dev.
Mitsubishi	98	3	0	3	2.42	0.641
Seat	98	3	0	3	1.98	0.786
Toyota	98	3	0	3	2.61	0.652
Total	98	8	1	9	7.01	1.388

Table A.4. Descriptive statistics for answers to structured questions. Mean value indicates mean number of questions answered correctly.

	n	Range	Min.	Max.	Mean	Std. Dev.
Mitsubishi	91	2	0	2	1.01	0.624
Seat	91	2	0	2	0.81	0.595
Toyota	91	3	1	4	2.18	0.769

Table A.5. Descriptive statistics for unstructured task. Mean value indicates mean number of models named for each brand (participants were asked to select four total models from among the three brands).

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
CAte1	0.199	0.840	0.180	0.082	0.115
CAte2	0.320	0.807	0.173	0.030	-0.007
CAte3	0.201	0.803	0.203	0.070	-0.184
CAim1	0.237	0.128	0.763	0.288	-0.113
CAim2	0.165	0.205	0.799	0.155	0.045
CAim3	0.199	0.264	0.704	-0.063	0.287
CAen1	0.801	0.146	0.193	0.212	0.102
CAen2	0.773	0.199	0.135	0.289	0.183
CAen3 R	0.746	0.162	0.134	0.096	0.300
CAco1	0.341	0.009	0.181	0.690	0.402
CAco2	0.259	0.122	0.171	0.843	0.105
CAco3 R	0.084	-0.081	0.077	0.246	0.879
CAcu1	0.744	0.385	0.181	0.027	-0.112
CAcu2	0.711	0.219	0.293	0.170	0.108
CAcu3	0.835	0.161	0.095	0.168	-0.166

Table A.6. Factor loadings, Cognitive Absorption. Rotated component matrix using Varimax with Kaiser normalization. CAte = Temporal Dissociation, CAim = Focused Immersion, CAen = Enjoyment, CAco = Control, CAcu = Curiosity. Note that "R" appearing at the end of an item designation indicates a reverse-coded item.

	Factor 1	Factor 2	Factor 3
INco1	0.471	0.308	0.604
INco2	0.156	-0.011	0.920
INco3 R	0.505	0.416	0.565
INtw1	0.188	0.803	0.055
INtw2	0.290	0.799	0.129
INtw3	0.215	0.814	0.121
INsy1	0.854	0.293	0.186
INsy2	0.854	0.259	0.185
INsy3	0.792	0.205	0.271

Table A.7. Factor loadings, Interactivity. Rotated component matrix using Varimax with Kaiser normalization. INco = Control, INtw = Two-Way Communication, INsy = Synchronicity. Note that "R" appearing at the end of an item designation indicates a reverse-coded item.

Construct	Dimension	Abbrev.	Item	Source	Adaptation(s)
CA	Temporal	cate1	Time appeared to go by very quickly when I used the site.	Agarwal & Karahanna (2000): "Time appears to go by very quickly when I am using the Web."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Temporal	cate2	Time flew when I used the site.	Agarwal & Karahanna (2000): "Time flies when I am using the Web."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Temporal	cate3	I lost track of time when using the site.	Agarwal & Karahanna (2000): "Sometimes I lose track of time when I am using the Web."	Adapted to single website (rather than "the Web"); eliminated "sometimes" and modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Immersion	caim1	While using the site, I was able to block out most other distractions.	Agarwal & Karahanna (2000): "While using the Web I am able to block out most other distractions."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Immersion	caim2	While on the site, my attention was not easily diverted.	Agarwal & Karahanna (2000): "While on the Web, my attentiondoes not get divertedvery easily."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Immersion	caim3	While on the site, I was immersed in the task I was performing.	Agarwal & Karahanna (2000): "While on the Web, I am immersed in the task I am performing."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Enjoyment	caen1	I had fun interacting with the site.	Agarwal & Karahanna (2000): "I have fun interacting with the Web."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Enjoyment	caen2	I enjoyed using the site.	Agarwal & Karahanna (2000): "I enjoy using the Web."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject

CA	Enjoyment	caen3	Using the site bored me.	Agarwal & Karahanna (2000): "Using the Web bores me."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Control	caco1	When using the site, I felt in control.	Agarwal & Karahanna (2000): "When using the Web I feel in control."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Control	caco2	The site allowed me to control my computer interaction.	Agarwal & Karahanna (2000): "The Web allows me to controlmy computerinteraction."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Control	caco3	I felt I had no control over my interaction with the site.	Agarwal & Karahanna (2000): "I feel I have no control over my interaction with the Web."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Curiosity	cacu1	Using the site excited my curiosity.	Agarwal & Karahanna (2000): "Using the Web excites my curiosity."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Curiosity	cacu2	Interacting with the site made me curious.	Agarwal & Karahanna (2000): "Interacting with the Web makes me curious."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
CA	Curiosity	cacu3	Using the site aroused my imagination.	Agarwal & Karahanna (2000): "Using the Web arouses my imagination."	Adapted to single website (rather than "the Web"); modified to past-tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
Interactivity	Control	inco1	While I was on the site, I could choose freely what I wanted to see.	Liu (2003): "While I was on the website, I could choose freely what I wanted to see.	"Website" changed to "site", consistent with Song & Zinkhan (2008) as well as other items within this same survey.
Interactivity	Control	inco2	While surfing the site, I had absolutely no control over what I could do on the site.	Liu (2003): "While surfing the website, I had absolutely no control over what I can do on the site."	"Can" changed to "could" and "website" changed to "site", consistent with Song & Zinkhan (2008) -- tense is otherwise inconsistent and past-tense indicates that we're asking about the one time the subject used the site as part of this experiment.

Interactivity	Control	inco3	I felt that I had a great deal of control over my visiting experience at the site.	Song & Zinkhan (2008): "I feel that I have a great deal of control over my visiting experience at this site."	Liu (2003): "I felt that I had a lot of control over my visiting experiences at this website." Changed tense to better indicate that we're asking about the one time the subject used the site as part of the experiment; changed "this" to "the" to be consistent with other items.
Interactivity	Two-Way Comm.	intw1	The site facilitated two-way communication between visitors and the site.	Liu (2003): "This website facilitates two-way communication between the visitors and the site."	"Website" changed to "site" consistent with Song & Zinkhan as well as other items within this same survey. Song & Zinkhan omitted "between the visitors and the site", which seems potentially confusing. Changed to past tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
Interactivity	Two-Way Comm.	intw2	The site made me feel it wants to listen to its visitors.	Liu (2003): "The website makes me feel it wants to listen to its visitors."	"Website" changed to "site" in order to be consistent with other items in this survey. Changed to past tense to indicate that we're asking about the one time the subject used the site as part of this experiment. Song & Zinkhan omitted this item without discussion, despite it being tied for the second-highest factor-loading item in Liu (and the highest among "traditional students" in that study).
Interactivity	Two-Way Comm.	intw3	The site is effective in gathering visitors' feedback.	Liu (2003): "The website is effective in gathering visitors' feedback."	"Website" changed to "site" for consistency. Retained "is", since we do not expect users to give the site feedback during their visits (as would be implied by using "was").
Interactivity	Synchronicity	insy1	Getting information from the site was very fast.	Liu (2003): "Getting information from the website is very fast."	"Website" changed to "site" in order to be consistent with other items in this survey. Changed to past tense to indicate that we're asking about the one time the subject used the site as part of this experiment.
Interactivity	Synchronicity	insy2	I was able to obtain the information I wanted without any delay.	Liu (2003): "I was able to obtain the information I want without any delay."	Verbatim.

Interactivity	Synchronicity	insy3	When I clicked on the links, I felt I was getting instantaneous information.	Liu (2003): "When I clicked on the links, I felt I was getting instantaneous information."	Verbatim.

Table A.8. Cognitive Absorption and Interactivity instrument items.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Time	1.000	0.705**	0.131	0.082	-0.152*	0.171*	0.011	-0.082	0.064	0.114	-0.063	-0.079	-0.061
2. Pages		1.000	0.151*	0.049	-0.158*	0.103	0.053	-0.004	0.117	0.120	0.076	0.044	0.076
3. Structure			1.000	0.111	-0.068	0.179*	0.003	-0.055	0.105	0.076	-0.039	-0.057	-0.242**
4. Cog. Abs.				1.000	0.406**	-0.001	-0.019	-0.009	0.040	0.129	0.126	0.029	0.133
5. Interact.					1.000	-0.047	-0.003	0.058	-0.100	0.157*	0.067	-0.039	0.046
6. Br. Fam.						1.000	0.019	0.004	0.009	-0.007	-0.098	0.028	0.044
7. Prod. Inv.							1.000	-0.018	-0.017	-0.008	0.064	0.082	0.050
8. Age								1.000	-0.032	-0.012	0.070	0.103	-0.007
9. Gender (F)									1.000	-0.010	-0.036	-0.030	-0.107
10. US Res.										1.000	-0.035	-0.094	0.072
11. Br. Att. Δ											1.000	0.517**	0.043
12. Purch. Δ												1.000	0.051
13. Know. Δ													1.000

Table A.9. Correlation matrix, regression variables. Correlation coefficients shown; * = significant at $\alpha = 0.05$, ** = significant at $\alpha = 0.01$.

	CA – Temp. Dissociation	CA – Immersion	CA – Enjoyment	CA – Control	CA – Curiosity	Inter. – Control	Inter. – Two-Way Comm.	Inter. - Synchronicity
CA – Temp.	1.000	0.496**	0.088	0.097	0.561**	0.112	0.299**	0.161*
CA – Imm.		1.000	0.093	0.150*	0.483**	0.077	0.292**	0.275**
CA – Enj.			1.000	0.535**	0.125	0.525**	0.092	0.079
CA – Cont.				1.000	0.169*	0.687**	0.131	0.138
CA – Curios.					1.000	0.212**	0.474**	0.342**
Int. – Cont.						1.000	0.147*	0.181*
Int. – 2-Way							1.000	0.556**
Int. – Synch.								1.000

Table A.10. Correlation matrix, sub-dimensions of Cognitive Absorption and Interactivity. Correlation coefficients shown; * = significant at $\alpha = 0.05$, ** = significant at $\alpha = 0.01$.

```xml
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<link rel="stylesheet" type="text/css" href="main.css"></link>
<title>Recorded Problem Steps</title>
</head>
<body>
<!-- This is the recorded XML problem data that was used in generating this page. -->
<xml id="recordeddata">
<?xml version="1.0" encoding="UTF-8"?>
<Report>
  <System MajorVersion="6" MinorVersion="1" ServicePackMajor="1" ServicePackMinor="0" BuildNumber="7601" Sku="4" Platform="1" /
  <UserActionData>
    <RecordSession SessionCount="1" StartTime="12:29:49 PM" StopTime="12:50:14 PM" ActionCount="382" MissedActionCount="0">
      <EachAction ActionNumber="1" Time="12:30:02 PM" Pid="3180" ProgramId="0000f519feec486de87ed73cb92d3cac802400000000" FileI
        <Description>User left click on "Mozilla Firefox (list item)" in "Program Manager"</Description>
        <Action>Mouse Left Click</Action>
        <CursorCoordsXY>41,751</CursorCoordsXY>
        <ScreenCoordsXYWH>0,0,2560,1024</ScreenCoordsXYWH>
        <UIStack>
          <Level Name="Mozilla Firefox" RoleId="34" Role="list item" Coordinates="691,0,74,85" />
          <Level Name="Desktop" RoleId="33" Role="list" Coordinates="0,0,2560,1024" />
          <Level Name="FolderView" RoleId="9" Role="window" ClassName="SysListView32" Coordinates="0,0,2560,1024" />
          <Level RoleId="9" Role="window" ClassName="SHELLDLL_DefView" Coordinates="0,0,2560,1024" />
          <Level Name="Program Manager" RoleId="9" Role="window" ClassName="Progman" Coordinates="0,0,2560,1024" />
        </UIStack>
      </EachAction>
      <EachAction ActionNumber="2" Time="12:30:07 PM" Pid="3180" ProgramId="0000f519feec486de87ed73cb92d3cac802400000000" FileI
        <Description>User left click on "Internet Explorer (list item)" in "Program Manager"</Description>
        <Action>Mouse Left Click</Action>
        <CursorCoordsXY>43,642</CursorCoordsXY>
        <ScreenCoordsXYWH>0,0,2560,1024</ScreenCoordsXYWH>
        <UIStack>
          <Level Name="Internet Explorer" RoleId="34" Role="list item" Coordinates="593,0,74,85" />
          <Level Name="Desktop" RoleId="33" Role="list" Coordinates="0,0,2560,1024" />
          <Level Name="FolderView" RoleId="9" Role="window" ClassName="SysListView32" Coordinates="0,0,2560,1024" />
          <Level RoleId="9" Role="window" ClassName="SHELLDLL_DefView" Coordinates="0,0,2560,1024" />
          <Level Name="Program Manager" RoleId="9" Role="window" ClassName="Progman" Coordinates="0,0,2560,1024" />
```

Figure A.1. Sample XML file created by Problem Steps Recorder.

```
208 12:27:00 PM 44820   Mouse Left Click    User left click on "X (editable text)" in "New Cars - Used Cars for Sale - SEAT | SEAT.co.uk - W
209 12:27:02 PM 44822   Mouse Left Click    User left click on "IBIZA SC (link)" in "New Cars - Used Cars for Sale - SEAT | SEAT.co.uk - Win
210 12:27:11 PM 44831   Mouse Left Click    User left click on "Survey | Qualtrics Survey Software | Current Progress 51% (page tab)" in "Ne
211 12:27:14 PM 44834   Mouse Left Click    User left click on "New SEAT Ibiza SC 2012 - Sport Coupé - Ibiza Sport Coupé | SEAT.co.uk (page
212 12:27:15 PM 44835   Mouse Wheel Down    User mouse wheel down on "BROCHURE REQUEST (editable text)" in "New SEAT Ibiza SC 2012 - Sport C
213 12:27:17 PM 44837   Mouse Wheel Up  User mouse wheel up on "New SEAT Ibiza SC 2012 - Sport Coupé - Ibiza Sport Coupé | SEAT.co.uk (pane)
214 12:27:20 PM 44840   Mouse Wheel Down    User mouse wheel down on "New SEAT Ibiza SC 2012 - Sport Coupé - Ibiza Sport Coupé | SEAT.co.uk
215 12:27:23 PM 44843   Mouse Wheel Up  User mouse wheel up on "New SEAT Ibiza SC 2012 - Sport Coupé - Ibiza Sport Coupé | SEAT.co.uk (pane)
216 12:27:31 PM 44851   Mouse Left Click    User left click on "VERSIONS (editable text)" in "New SEAT Ibiza SC 2012 - Sport Coupé - Ibiza S
217 12:27:35 PM 44855   Mouse Left Click    User left click on "New SEAT Ibiza SC 2012 - Sport Coupé Versions | SEAT.co.uk (pane)" in "New S
218 12:27:36 PM 44856   Mouse Left Click    User left click on "FR (editable text)" in "New SEAT Ibiza SC 2012 - Sport Coupé Versions | SEAT
219 12:27:37 PM 44857   Mouse Wheel Down    User mouse wheel down on "SEAT Ibiza - SC FR | SEAT.co.uk (pane)" in "SEAT Ibiza - SC FR | SEAT.
220 12:27:40 PM 44860   Mouse Wheel Up  User mouse wheel up on "FR Interior Dash (graphic)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Windows In
221 12:27:46 PM 44866   Mouse Left Double Click User left double click on "You'll love the SEAT Ibiza FR. It's designed to be driven - with
222 12:27:46 PM 44866   Mouse Left Click    User left click on "You'll love the SEAT Ibiza FR. It's designed to be driven - with great looks
223 12:27:47 PM 44867   Mouse Left Click    User left click on "SEAT Ibiza - SC FR | SEAT.co.uk (pane)" in "SEAT Ibiza - SC FR | SEAT.co.uk
224 12:27:50 PM 44870   Mouse Wheel Down    User mouse wheel down on "SEAT Ibiza - SC FR | SEAT.co.uk (pane)" in "SEAT Ibiza - SC FR | SEAT.
225 12:27:50 PM 44870   Mouse Wheel Up  User mouse wheel up on "The sleek design of the SEAT Ibiza FR includes headlights with optional LEDs
226 12:27:51 PM 44871   Mouse Wheel Down    User mouse wheel down on "SEAT Ibiza - SC FR | SEAT.co.uk (pane)" in "SEAT Ibiza - SC FR | SEAT.
227 12:27:55 PM 44875   Mouse Left Click    User left click on "Next (link)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Windows Internet Explorer
228 12:27:56 PM 44876   Mouse Left Click    User left click on "Next (link)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Windows Internet Explorer
229 12:27:57 PM 44877   Mouse Left Click    User left click on "Next (link)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Windows Internet Explorer
230 12:27:58 PM 44878   Mouse Wheel Up  User mouse wheel up on "SEAT Ibiza - SC FR | SEAT.co.uk (pane)" in "SEAT Ibiza - SC FR | SEAT.co.uk
231 12:27:58 PM 44878   Mouse Wheel Down    User mouse wheel down on "Diesel Engines  (editable text)" in "SEAT Ibiza - SC FR | SEAT.co.uk -
232 12:28:02 PM 44882   Mouse Left Click    User left click on "Car Configurator CTA (graphic)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Window
233 12:28:08 PM 44888   Mouse Wheel Down    User mouse wheel down in "SEAT Car Configurator - Windows Internet Explorer"
234 12:28:12 PM 44892   Mouse Left Click    User left click on "Continue to Engine (link)" in "SEAT Car Configurator - Windows Internet Expl
235 12:28:15 PM 44895   Mouse Left Click    User left click on "Continue to Design (link)" in "SEAT Car Configurator - Windows Internet Expl
236 12:28:34 PM 44914   Mouse Left Click    User left click on "Survey | Qualtrics Survey Software | Current Progress 51% (page tab)" in "SE
237 12:28:35 PM 44915   Mouse Left Click    User left click on "9 (radio button)" in "Survey | Qualtrics Survey Software | Current Progress
238 12:28:36 PM 44916   Mouse Wheel Down    User mouse wheel down on "Survey | Qualtrics Survey Software | Current Progress 51% (pane)" in "
239 12:28:42 PM 44922   Mouse Left Click    User left click in "Survey | Qualtrics Survey Software | Current Progress 51% - Windows Internet
240 12:28:43 PM 44923   Mouse Left Click    User left click on "Cars, Crossovers, Coupes, Convertibles, Sedans &amp; Electric / Mitsubishi M
241 12:28:44 PM 44924   Mouse Left Click    User left click on "SEAT Ibiza - SC FR | SEAT.co.uk (page tab)" in "Cars, Crossovers, Coupes, Co
242 12:28:45 PM 44925   Mouse Left Click    User left click on "SEAT Car Configurator (page tab)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Wind
243 12:28:47 PM 44927   Mouse Left Click    User left click on "Close Tab (Ctrl+W) (push button)" in "SEAT Car Configurator - Windows Intern
244 12:28:48 PM 44928   Mouse Left Click    User left click on "SEAT Ibiza - SC FR | SEAT.co.uk (page tab)" in "Cars, Crossovers, Coupes, Co
245 12:28:49 PM 44929   Mouse Wheel Up  User mouse wheel up on "The Ibiza FR is bursting with technology. A latest generation electronic XDS
246 12:28:51 PM 44931   Mouse Left Click    User left click on "ALTEA XL (link)" in "SEAT Ibiza - SC FR | SEAT.co.uk - Windows Internet Expl
```

Figure A.2. Sample of tab-delimited file parsed from XML file in Figure A.1.

Appendix B

Criterion	Description	Coder 1	Coder 2	Coder 3	Mean
11 Minutes	An 11-minute gap between clicks signals a break between sessions (Göker & He, 2000).	71.7%	71.0%	68.9%	70.5%
24 Hours	A 24-hour gap between clicks signals a break between sessions.	55.4%	55.3%	55.3%	55.3%
Mean + 2σ	A gap two sigmas or more beyond the mean gap signals a break between sessions.	27.5%	27.5%	27.4%	27.5%
Mean + 3σ	A gap three sigmas or more beyond the mean gap signals a break between sessions.	40.5%	40.3%	39.6%	40.1%
Median + 2 MADs	A gap two median absolute deviations from the median gap signals a break between sessions.	46.9%	46.7%	46.1%	46.6%
Median + 3 MADs	A gap three median absolute deviations from the median gap signals a break between sessions.	3.4%	3.4%	3.4%	3.4%

Table B.1. Comparison of candidate session identification methods.

Variable	*n*	Mean	S.D.	Min.	.25	Med.	.75	Max.
Purchase	1,823,539	0.01	0.10	0	0	0	0	1
Number of Purchases	1,823,539	0.08	0.36	0	0	0	0	20
Location Breadth	1,823,539	0.01	0.087	0	0	0	0	8
Other Breadth	1,823,539	0.01	0.108	0	0	0	0	9
Airline Name	1,823,539	0.04	0.202	0	0	0	0	1
Brand Name	1,823,539	0.07	0.261	0	0	0	0	1
Location Depth	1,823,539	1.27	1.476	0	0	0	3	4
Other Depth	1,823,539	1.02	0.762	0	0	1	2	4
Number of Queries	1,823,539	1.12	0.47	1	1	1	1	38
Session Duration	1,823,539	0.39	1.77	0	0	0	0	68

Table B.2. Descriptive statistics for regression model variables.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. Purchase	1.000	0.675**	-0.005**	-0.002**	0.003**	0.033**	-0.003**	0.027**	0.493**	0.402**
2. Number of Purchases		1.000	-0.014**	-0.012**	0.004**	0.077**	-0.09**	0.073**	0.577**	0.393**
3. Location Breadth			1.000	0.427**	0.000	-0.022**	0.239**	-0.061**	0.201**	0.253**
4. Other Breadth				1.000	-0.001	-0.014**	0.432**	-0.039**	0.253**	0.317**
5. Airline Name					1.000	-0.059**	0.003**	-0.185**	0.004**	-0.002**
6. Brand Name						1.000	-0.011**	0.201**	0.036**	0.020**
7. Location Depth							1.000	-0.026**	0.161**	0.221**
8. Other Depth								1.000	0.017**	0.011**
9. Number of Queries									1.000	0.756**
10. Session Duration										1.000

Table B.3. Correlation matrix for regression model variables. Correlation coefficients shown; * = significant at $\alpha = 0.05$, ** = significant at $\alpha = 0.01$.

Variable	B	SE	p	Exp(B)	Supported?
Constant	-5.938				
Airline Name	0.372	0.042	< 0.001	1.451	H2 Supported
Brand Name	0.619	0.025	< 0.001	1.857	H2 Supported
Location Depth	-0.053	0.007	< 0.001	0.949	H2 Not Supported
Other Depth	0.307	0.014	< 0.001	1.359	H2 Supported
Session Duration	0.416	0.002	< 0.001	1.515	

Table B.4. Initial logistic regression omitting breadth variables, full data set.

Variable	B	SE	p	Exp(B)	Supported?
Constant	-5.979				
Location Breadth	-3.992	0.274	< 0.001	0.018	H1 Supported
Other Breadth	-3.857	0.175	< 0.001	0.021	H1 Supported
Airline Name	0.405	0.048	< 0.001	1.499	H2 Supported
Brand Name	0.642	0.028	< 0.001	1.900	H2 Supported
Location Depth	-0.008	0.008	0.358	0.992	H2 Not Supported
Other Depth	0.276	0.015	< 0.001	1.317	H2 Supported
Session Duration	0.457	0.002	< 0.001	1.580	

Table B.5. Breadth and specificity test results using rare events logistic regression, full data set.

Variable	*n*	Mean	S.D.	Min.	.25	Med.	.75	Max.
Purchase	171,008	0.10	0.304	0	0	0	0	1
Location Breadth	171,008	0.07	0.274	0	0	0	0	8
Other Breadth	171,008	0.11	0.324	0	0	0	0	9
Airline Name	171,008	0.05	0.217	0	0	0	0	1
Brand Name	171,008	0.11	0.309	0	0	0	0	1
Location Depth	171,008	1.09	1.412	0	0	0	2	4
Other Depth	171,008	1.05	0.754	0	0	1	2	3
Number of Queries	171,008	2.30	0.901	2	2	2	2	38
Session Duration	171,008	4.14	4.234	0	0	3	7	68

Table B.6. Descriptive statistics for regression model variables, multiple-query sessions. Session duration given in minutes from first query to last.

Variable	*n*	Mean	S.D.	Min.	.25	Med.	.75	Max.
Purchase	1,652,531	4.24×10^{-6}	0.002	0	0	0	0	1
Location Breadth	1,652,531	0		0	0	0	0	0
Other Breadth	1,652,531	1.21×10^{-4}	0.008	0	0	0	0	1.5
Airline Name	1,652,531	0.04	0.200	0	0	0	0	1
Brand Name	1,652,531	0.07	0.256	0	0	0	0	1
Location Depth	1,652,531	1.29	1.481	0	0	0	3	4
Other Depth	1,652,531	1.01	0.762	0	0	1	2	4
Number of Queries	1,652,231	1		1	1	1	1	1
Session Duration	1,652,231	0		0	0	0	0	0

Table B.7. Descriptive statistics for regression model variables, single-query sessions. All values of Location Breadth and Session Duration for this sub-sample were 0; all values for Number of Queries were 1.

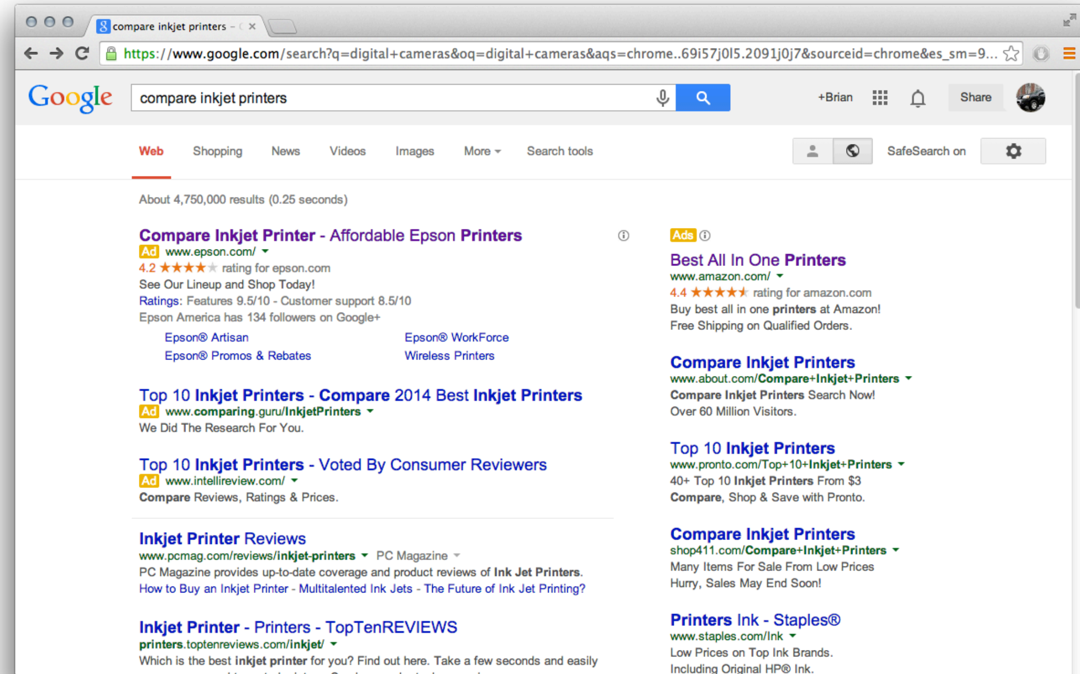


Figure B.1. Sample search results page. Sponsored search advertisements can be found on the top of the page and on the right side.

Appendix C

Candidate Story	*n*	Mean	S.D.	Min.	.25	Med.	.75	Max.
The recent election in Norway	21	2.24	1.37	1	1	2	3	5
Political unrest in Egypt	21	5.00	1.22	3	4	5	6	7
Attack on citizens at a mall in Nairobi, Kenya	21	3.71	2.31	1	1	4	5	7
The initial public stock offering of Twitter	21	4.48	2.04	1	3	5	6	7
Changes in US leading economic indicators	21	5.19	2.14	1	4	6	7	7
The competition between Microsoft's new Xbox One and Sony's PlayStation4	21	3.95	1.96	1	3	4	6	7
New capabilities being added to Samsung's Galaxy smartphones	20	4.75	1.94	1	3.5	5	6.5	7
Disagreement about universal healthcare (and/or "Obamacare") in the US	21	4.33	1.93	1	3	4	6	7
The US government's acquisition and use of individuals' phone records	21	5.10	1.61	2	4	5	6	7
Group of motorcyclists beating up a motorist in New York	21	3.33	1.77	1	2	4	4	7
Crash of a hot air balloon in New Mexico	21	2.10	1.61	1	1	1	3	6
Kanye West's recent comments on talk shows	21	3.52	2.09	1	2	3	5	7
Movies scheduled for release during the upcoming holiday season	21	4.95	1.75	1	4	5	6	7
Breakthrough in treatment of Alzheimers disease	21	4.86	1.82	1	4	5	7	7
Link between genetics and marital happiness	21	4.24	1.73	1	3	4	5	7

Table C.1. Descriptive statistics for task subject matter candidates. Data taken from pilot study.

Internet News Use	inu1	I use the Internet to find information about news stories frequently.
Internet News Use	inu2	The Internet is a good source for timely information about national and world events.
Internet News Use	inu3	If I want to find out about a news story, I use the Internet.
Task Subject Knowledge	tsk1	I know a lot about Norway.
Task Subject Knowledge	tsk2	I'm likely to be interested in a news story if it involves Norway.
Task Subject Knowledge	tsk3	I keep up-to-date with important events in Norway

Table C.2. Survey items for Pre-Task Stage instruments.

	inu1	inu2	inu3	tsk1	tsk2	tsk3
inu1	1.000	0.623**	0.666**	0.011	0.104	0.029
inu2		1.000	0.786**	-0.080	0.093	-0.070
inu3			1.000	-0.100	0.053	-0.100
tsk1				1.000	0.392**	0.644**
tsk2					1.000	0.483**
						1.000

Table C.3. Item correlation matrix, Internet News Use and Task Subject Knowledge. * = significant at α = 0.05, ** = significant at α = 0.01.

Variable	*n*	Mean	S.D.	Min.	.25	Med.	.75	Max.
inu1	241	6.03	1.14	1	6	6	7	7
inu2	241	6.31	0.89	1	6	7	7	7
inu3	241	6.27	0.86	1	6	6	7	7
tsk1	241	2.20	1.17	1	1	2	3	5
tsk2	241	3.33	1.57	1	2	3	5	7
tsk3	241	2.08	1.07	1	1	2	3	6

Table C.4. Descriptive statistics for Internet News Use and Task Subject Knowledge items.

Article	Headline	Words	Grade	Answers
A	"Center-Right Alliance Victorious in Norway Elections"	506	11.8	
B	"Pirate Party, Others Fail to Breach Norwegian Parliament"	577	11.7	UC5
C	"Anti-Immigrant Party Linked to Mass Murderer Set to Enter Government"	655	11.9	UC4
D	"Norwegian Election: Conservative Coalition Triumphant"	505	12.1	
E	"Conservative Party Sweeps into Power in Norwegian Elections"	767	12.2	
F	"Norway Risks Economic Overheating as Opposition Scores Victory"	727	12.5	
G	"Stoltenberg: Sort of an Every-Man, No Longer PM"	623	12.5	UC3
H	"Norway Shifts Right in Post-Breivik Election"	741	11.1	UC2
I	"Rich Norwegians Turn Down Labor"	640	11.0	UC1

Table C.5. Article headlines and statistics. "Grade" reflects Fleisch-Kincade grade level score.

ID	Question	% Correct
C1	What is the name of the person who will have become prime minister as a result of the elections?	97.6%
C2	Following the recent Norwegian election, with which other party was the Conservative Party expected to form a coalition?	76.6%
C3	What proportion of voters voted for the Labor Party in the election?	97.6%
C4	When was the election held?	92.7%
C5	How many parties would be represented in parliament as a result of receiving a high enough proportion of the votes?	71.8%
UC1	What is the home city of the incoming prime minister?	46.8%
UC2	What was the approximate official unemployment rate for Norway in September 2013?	74.2%
UC3	What occupation did the departing prime minster, Jens Stoltenberg, perform for a day as part of his campaign?	80.6%
UC4	How many deaths occurred in the 2011 terror bombings carried out by Anders Breivik?	88.7%
UC5	How many people must sign a party's petition for that party to be included in the national elections?	64.5%

Table C.6. High structure task questions. Answers to common-type questions could be found in any of the nine articles; answers to the uncommon-type questions could only be found in one of the articles each.

ID	Question Text
1	If you were a Norwegian considering whom to vote for in the 2013 election, what would have been the most important political issues involved in your decision?
2	Explain why each of those issues would have been important to you.
3	Choose one of the issues you mentioned above and interpret the debate that the parties had around that topic during the election.
4	If you had been a Norwegian voting in the election, for which party would you have voted?
5	Why would you have voted for that party?

Table C.7. Low structure task questions.

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
cate1	0.3856	0.7655	0.1641	0.0885	0.0009
cate2	0.3369	0.8152	0.0942	0.1417	-0.0242
cate3	0.3041	0.6423	0.0335	0.1487	0.2097
caim1	0.1072	0.1536	0.1535	0.6241	0.0412
caim2	0.1963	0.1423	0.2444	0.6485	0.1055
caim3	0.3069	0.3064	0.1693	0.5238	-0.1911
caen1	0.7713	0.3106	0.2497	0.1322	-0.0029
caen2	0.7651	0.2941	0.3134	0.0922	-0.0152
caen3	0.7402	0.2334	0.1470	0.1269	-0.1698
caco1	0.3106	0.1557	0.7326	0.1920	0.0270
caco2	0.3361	0.1479	0.6315	0.1764	0.1318
caco3	0.5311	0.3743	0.1049	0.1493	0.3699
cacu1	0.6999	0.3732	0.1771	0.1238	0.2026
cacu2	0.7101	0.3325	0.0821	0.1712	0.1824
cacu3	0.5311	0.3743	0.1049	0.1493	0.3699

Table C.8. Confirmatory factor analysis for Cognitive Absorption. Varimax rotation applied.

Item	Factor 1	Factor 2
siin1	0.6846	0.1270
siin2	0.7934	0.2123
siin3	0.8332	0.2306
siin4	0.8594	0.1857
siin5	0.6250	0.4230
siin6	0.7649	0.3091
siin7	0.7532	0.2359
siin8	0.7024	0.3337
siii1	0.5101	0.4551
siii2	0.1957	0.6770
siii3	0.3348	0.7081
siii4	0.1987	0.7651

Table C.9. Confirmatory factor analysis for Susceptibility to Interpersonal Influence. Varimax rotation applied. In the above table, "siin" refers to the normative sub-dimension of Susceptibility to Interpersonal Influence, while "siii" refers to the informational sub-dimension.

	fam1	fam2	fam3	freq
fam1	1.000	0.794**	0.690**	0.829**
fam2		1.000	0.829**	0.831**
fam3			1.000	0.861**
freq				1.000

Table C.10. Correlation matrix, familiarity and frequency of use items. * = significant at $\alpha = 0.05$, ** = significant at $\alpha = 0.01$.

	n	**Mean**	**S.D.**	**Min.**	**.25**	**Med.**	**.75**	**Max.**
BBC News								
Familiarity	241	3.96	1.89	1.00	2.00	4.00	5.33	7.00
Frequency of Use	241	3.52	1.89	1.00	2.00	3.00	5.00	7.00
DW.de								
Familiarity	241	1.38	077	1.00	1.00	1.00	1.67	6.33
Frequency of Use	241	1.12	0.55	1.00	1.00	1.00	1.00	6.00
The Economist								
Familiarity	241	2.65	1.60	1.00	1.00	2.00	4.00	6.67
Frequency of Use	241	2.15	1.45	1.00	1.00	2.00	3.00	7.00
The Independent								
Familiarity	241	1.82	1.23	1.00	1.00	1.00	2.00	6.33
Frequency of Use	241	1.59	1.12	1.00	1.00	1.00	2.00	7.00
Inter Press Service								
Familiarity	241	1.41	0.76	1.00	1.00	1.00	2.00	5.00
Frequency of Use	241	1.19	0.67	1.00	1.00	1.00	1.00	5.00
LATimes.com								
Familiarity	241	3.53	1.72	1.00	2.00	3.67	5.00	7.00
Frequency of Use	241	2.86	1.67	1.00	1.00	3.00	4.00	7.00
NBCNews								
Familiarity	241	4.30	1.72	1.00	3.00	4.67	5.67	7.00
Frequency of Use	241	3.64	1.81	1.00	2.00	4.00	5.00	7.00
Reuters								
Familiarity	241	3.46	1.94	1.00	1.67	3.67	5.00	7.00
Frequency of Use	241	3.00	1.81	1.00	1.00	3.00	5.00	7.00

Table C.11. Descriptive statistics for familiarity and frequency of use by website.

	n	1	2	3	4	5	6
Structured	124	58.99	984.01	4.38	2.82	6.22	2.45
High Borders	42	57.05	904.98	4.47	2.97	6.31	2.48
Medium Borders	39	57.67	1054.49	4.38	2.70	6.13	2.43
Low Borders	43	62.09	997.26	4.29	2.77	6.21	2.44
Unstructured	117	19.80	931.79	4.57	2.60	6.19	2.63
High Borders	41	18.29	862.57	4.79	2.80	6.26	2.91
Medium Borders	38	19.74	938.29	4.18	2.35	6.09	2.32
Low Borders	38	21.50	999.99	4.73	2.64	6.22	2.62

Table C.12. Comparison of user-level variable means by condition. 1 = total of site interactions, 2 = task completion time (in seconds), 3 = cognitive absorption (curiosity), 4 = susceptibility to interpersonal influence (SII; normative), 5 = internet news use, 6 = task subject knowledge.

	n	**Mean**	**S.D.**	**Min.**	**.25**	**Med.**	**.75**	**Max.**
1. Interact	241	39.97	35.67	0	15	28	56	223
2. Interstitial	241	4.98	8.35	0	0	1	7	54
3. Task Time	241	958.66	476.55	30.10	627.86	896.38	1201.17	2747.01
4. CA	241	4.47	1.37	1.00	3.67	4.67	5.33	7.00
5. SII	241	2.71	1.34	1.00	1.62	2.38	3.62	7.00
6. Internet News	241	6.20	0.86	1.00	6.00	6.33	7.00	7.00
7. Task Subj. Know.	241	2.54	1.03	1.00	2.00	2.33	3.00	6.00

Table C.13. Descriptive statistics, user-level variables.

Site	1	2	3	4	5	6	7	8
BBC News	3.95	3.51	4.65	69.3%	70.1%	88.9%	9.0%	32.3%
DW.de	1.37	1.12	5.16	63.9%	47.3%	86.8%	35.7%	22.2%
The Economist	2.65	2.14	4.90	66.0%	66.0%	81.1%	18.9%	30.9%
The Independent	1.82	1.59	4.90	71.0%	53.1%	87.5%	34.5%	28.4%
IPS	1.41	1.18	5.16	66.0%	45.6%	91.8%	36.5%	14.1%
LATimes.com	3.53	2.86	5.01	69.3%	56.8%	89.8%	26.3%	30.2%
Reuters	3.46	2.99	5.05	66.4%	62.7%	84.8%	20.0%	23.4%
The Nordic Page	1.32	1.08	5.18	68.0%	60.2%	90.3%	20.1%	29.1%
NBCNews.com	4.28	3.63	4.80	66.0%	56.4%	83.8%	28.3%	29.6%

Table C.14. Comparison of site-level variable means by site. 1 = familiarity, 2 = frequency of use, 3 = rank position on page, 4 = visited percentage, 5 = recognition percentage, 6 = percent of recognitions correct, 7 = percent of visits unrecognized, 8 = average attribution.

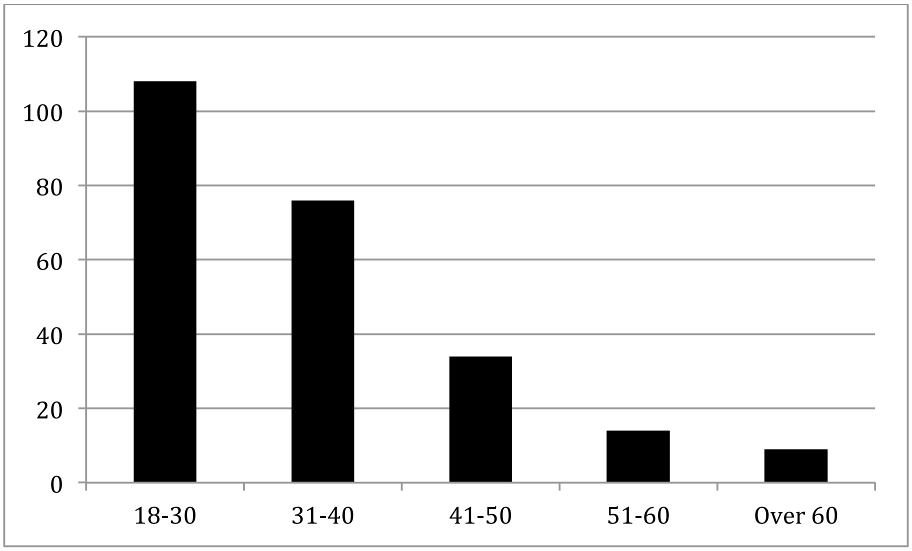


Figure C.1. User sample distribution by age range.

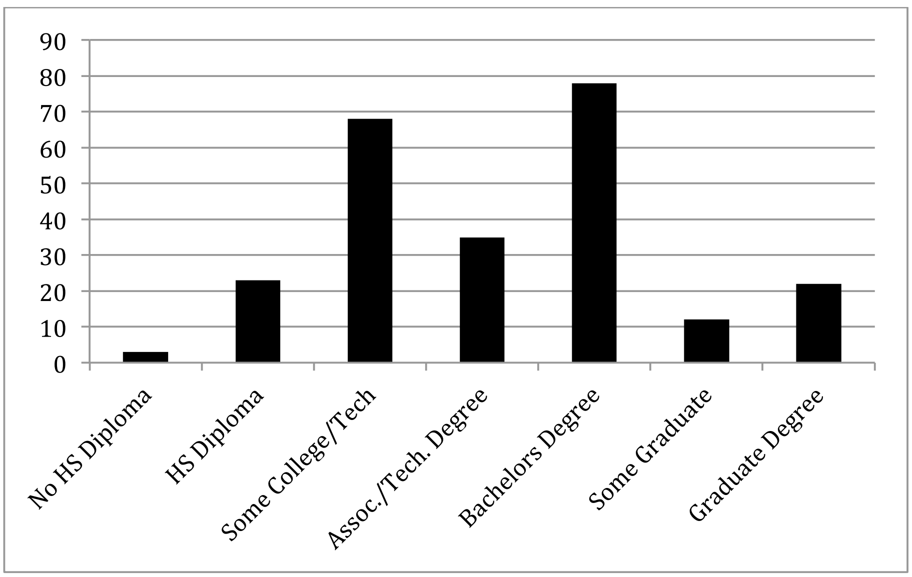


Figure C.2. User sample distribution by highest education level attained.

Google Norway Elections

News Search Results for: "Norway Elections"

Page 1 of 1

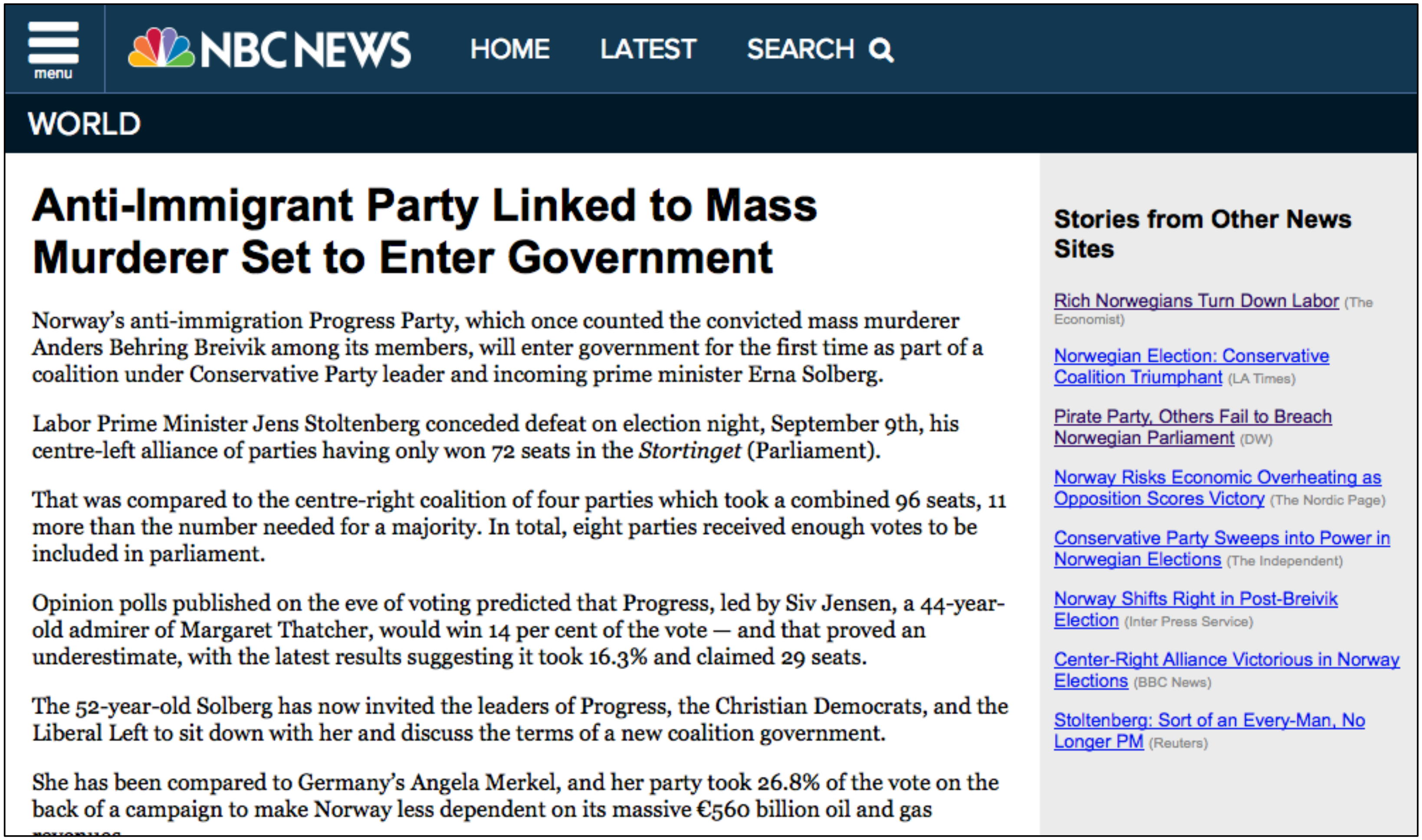
Rich Norwegians Turn Down Labor
The Economist
Promises of more welfare for Scandinavia's richest citizens failed to boost Labor Prime Minister Jens Stoltenberg's popularity. Voters instead handed power to a Conservative-led opposition block that promised tax cuts and more infrastructure investments.

Norwegian Election: Conservative Coalition Triumphant
LA Times
Norwegian center-right parties are set to form a new government after Labor Party leader Jens Stoltenberg admitted defeat. Result described as "a historic election victory for the right-wing parties".

Pirate Party, Others Fail to Breach Norwegian Parliament
DW
While the Conservative Party were the clear winners in Norway's election, a number of smaller parties had their hopes of winning enough votes to enter parliament summarily dashed.

Norway Risks Economic Overheating as Opposition Scores Victory
The Nordic Page
With Norway's opposition Conservatives sweeping into power in recent elections, analysts warn that the country's economy is at risk of overheating due to rising oil prices and a debt-fueled housing boom.

Conservative Party Sweeps into Power in Norwegian Elections
The Independent
Norway's opposition Conservatives, promising tax cuts and better healthcare, won elections in a landslide on Monday, but faced tough coalition talks with a populist party that wants to spend more of the country's



menu NBC NEWS HOME LATEST SEARCH

WORLD

Anti-Immigrant Party Linked to Mass Murderer Set to Enter Government

Norway's anti-immigration Progress Party, which once counted the convicted mass murderer Anders Behring Breivik among its members, will enter government for the first time as part of a coalition under Conservative Party leader and incoming prime minister Erna Solberg.

Labor Prime Minister Jens Stoltenberg conceded defeat on election night, September 9th, his centre-left alliance of parties having only won 72 seats in the *Stortinget* (Parliament).

That was compared to the centre-right coalition of four parties which took a combined 96 seats, 11 more than the number needed for a majority. In total, eight parties received enough votes to be included in parliament.

Opinion polls published on the eve of voting predicted that Progress, led by Siv Jensen, a 44-year-old admirer of Margaret Thatcher, would win 14 per cent of the vote — and that proved an underestimate, with the latest results suggesting it took 16.3% and claimed 29 seats.

The 52-year-old Solberg has now invited the leaders of Progress, the Christian Democrats, and the Liberal Left to sit down with her and discuss the terms of a new coalition government.

She has been compared to Germany's Angela Merkel, and her party took 26.8% of the vote on the back of a campaign to make Norway less dependent on its massive €560 billion oil and gas

Stories from Other News Sites

Rich Norwegians Turn Down Labor (The Economist)

Norwegian Election: Conservative Coalition Triumphant (LA Times)

Pirate Party, Others Fail to Breach Norwegian Parliament (DW)

Norway Risks Economic Overheating as Opposition Scores Victory (The Nordic Page)

Conservative Party Sweeps into Power in Norwegian Elections (The Independent)

Norway Shifts Right in Post-Breivik Election (Inter Press Service)

Center-Right Alliance Victorious in Norway Elections (BBC News)

Stoltenberg: Sort of an Every-Man, No Longer PM (Reuters)

Figure C.4. Sample branded content page.

Figure C.5. Sample interstitial pages for Medium (above) and High (below) border conditions.

```
</table>

<table cellpadding="0" cellspacing="0" border="0" name="ProblemStep15" ID="ProblemStep15" title="User mouse wheel down in &quot;Pirate Party, Others Fail to
<tr>
  <td class="contentcell-width stepOptionsDivider">
    <div class="topdivider"></div><a title="Previous" href="#ProblemStep14">
    <id ID="PreviousStepL">Previous</id></a>
    <a title="Next" href="#ProblemStep16">
    <id ID="NextStepL">Next</id></a>
  </td>
</tr>
<tr>
  <td class="contentcell-width">
    <p ID="ProblemStepP"><b>Problem Step 15: (4/23/2014 7:22:28 AM)</b>
    User mouse wheel down in &quot;Pirate Party, Others Fail to Breach Norwegian Parliament - NBC News.com - Google Chrome&quot;
    </p>
  </td>
</tr>
<tr>
  <td class="contentcell-width">
    <div class="contentcell-width noscreenshots"><p><i>No screenshots were saved for this problem step.</i></p></div>
  </td>
</tr>
<tr>
  <td class="contentcell-width titleSpacingAndDivider"> </td>
</tr>
</table>

<table cellpadding="0" cellspacing="0" border="0" name="ProblemStep16" ID="ProblemStep16" title="User mouse wheel up in &quot;Pirate Party, Others Fail to Br
<tr>
  <td class="contentcell-width stepOptionsDivider">
    <div class="topdivider"></div><a title="Previous" href="#ProblemStep15">
    <id ID="PreviousStepL">Previous</id></a>
    <a title="Next" href="#ProblemStep17">
    <id ID="NextStepL">Next</id></a>
  </td>
</tr>
<tr>
  <td class="contentcell-width">
    <p ID="ProblemStepP"><b>Problem Step 16: (4/23/2014 7:23:55 AM)</b>
    User mouse wheel up in &quot;Pirate Party, Others Fail to Breach Norwegian Parliament - NBC News.com - Google Chrome&quot;
    </p>
  </td>
</tr>
<tr>
  <td class="contentcell-width">
    <div class="contentcell-width noscreenshots"><p><i>No screenshots were saved for this problem step.</i></p></div>
  </td>
```


Figure C.6. Sample XML file created by Problem Steps Recorder.

```
1	7:19:07 AM	26347	Mouse Wheel Down	User mouse wheel down in "Survey | Qualtrics Survey Software | Current Progress 3% - Google Chrome"
2	7:19:12 AM	26352	Mouse Left Click	User left click in "Survey | Qualtrics Survey Software | Current Progress 3% - Google Chrome"
3	7:19:14 AM	26354	Keyboard Input	User keyboard input in "Survey | Qualtrics Survey Software | Current Progress 3% - Google Chrome"
4	7:19:17 AM	26357	Mouse Wheel Down	User mouse wheel down in "Survey | Qualtrics Survey Software | Current Progress 3% - Google Chrome"
5	7:19:20 AM	26360	Mouse Left Click	User left click in "Survey | Qualtrics Survey Software | Current Progress 3% - Google Chrome"
6	7:19:40 AM	26380	Mouse Wheel Down	User mouse wheel down in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
7	7:19:41 AM	26381	Mouse Right Click	User right click in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
8	7:19:42 AM	26382	Mouse Left Click	User left click on "Open link in new tab (menu item)"
9	7:19:44 AM	26384	Mouse Wheel Down	User mouse wheel down in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
10	7:20:07 AM	26407	Mouse Left Click	User left click on "News Search Results: 'Norway Election' (page tab)" in "Survey | Qualtrics Survey Software | Curre
11	7:20:57 AM	26457	Mouse Right Click	User right clicked
12	7:20:58 AM	26458	Mouse Left Click	User left click on "Open link in new tab (menu item)"
13	7:20:59 AM	26459	Mouse Wheel Down	User mouse wheel down
14	7:21:12 AM	26472	Mouse Left Click	User left click on "Redirecting to NBC News (page tab)"
15	7:22:28 AM	26548	Mouse Wheel Down	User mouse wheel down in "Pirate Party, Others Fail to Breach Norwegian Parliament - NBC News.com - Google Chrome"
16	7:23:55 AM	26635	Mouse Wheel Up	User mouse wheel up in "Pirate Party, Others Fail to Breach Norwegian Parliament - NBC News.com - Google Chrome"
17	7:23:56 AM	26636	Mouse Left Click	User left click on "Close (push button)" in "Pirate Party, Others Fail to Breach Norwegian Parliament - NBC News.com
18	7:23:59 AM	26639	Mouse Wheel Down	User mouse wheel down
19	7:24:01 AM	26641	Mouse Wheel Up	User mouse wheel up
20	7:24:09 AM	26649	Mouse Right Click	User right clicked
21	7:24:10 AM	26650	Mouse Left Click	User left click on "Open link in new tab (menu item)"
22	7:24:18 AM	26658	Mouse Wheel Down	User mouse wheel down
23	7:24:25 AM	26665	Mouse Left Click	User left click on "Anti-Immigrant Party Linked to Mass Murderer Set to Enter Government - The Nordic Page - Politics
24	7:24:32 AM	26672	Mouse Wheel Down	User mouse wheel down in "Anti-Immigrant Party Linked to Mass Murderer Set to Enter Government - The Nordic Page - Po
25	7:24:35 AM	26675	Mouse Left Click	User left click on "Close (push button)" in "Anti-Immigrant Party Linked to Mass Murderer Set to Enter Government - T
26	7:24:38 AM	26678	Mouse Right Click	User right clicked
27	7:24:39 AM	26679	Mouse Left Click	User left click on "Open link in new tab (menu item)"
28	7:24:40 AM	26680	Mouse Right Click	User right clicked
29	7:24:40 AM	26680	Mouse Left Click	User left click on "Open link in new tab (menu item)"
30	7:24:41 AM	26681	Mouse Right Click	User right clicked
31	7:24:42 AM	26682	Mouse Left Click	User left click on "Open link in new tab (menu item)"
32	7:24:43 AM	26683	Mouse Right Click	User right clicked
33	7:24:44 AM	26684	Mouse Left Click	User left click on "Open link in new tab (menu item)"
34	7:24:45 AM	26685	Mouse Left Click	User left clicked
35	7:24:51 AM	26691	Mouse Left Click	User left click on "Close (push button)" in "Redirecting to Inter Press Service - Google Chrome"
36	7:24:56 AM	26696	Mouse Left Click	User left click in "Redirecting to The Economist - Google Chrome"
37	7:25:39 AM	26739	Mouse Wheel Down	User mouse wheel down in "Rich Norwegians Turn Down Labor | The Economist - Google Chrome"
38	7:25:58 AM	26758	Mouse Left Click	User left click on "Close (push button)" in "Rich Norwegians Turn Down Labor | The Economist - Google Chrome"
39	7:26:05 AM	26765	Mouse Left Click	User left click on "Close (push button)" in "BBC News - Stoltenberg: Sort of an Every-Man, No Longer PM - Google Chro
40	7:26:10 AM	26770	Mouse Wheel Down	User mouse wheel down in "Conservative Party Sweeps into Power in Norwegian Elections | Reuters - Google Chrome"
41	7:26:13 AM	26773	Mouse Left Click	User left click on "Close (push button)" in "Conservative Party Sweeps into Power in Norwegian Elections | Reuters -
42	7:26:15 AM	26775	Mouse Left Click	User left click on "Survey | Qualtrics Survey Software | Current Progress 42% (page tab)"
43	7:26:17 AM	26777	Mouse Left Click	User left click in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
44	7:26:24 AM	26784	Keyboard Input	User keyboard input in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
45	7:27:14 AM	26834	Mouse Wheel Down	User mouse wheel down in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
46	7:27:17 AM	26837	Keyboard Input	User keyboard input in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
47	7:27:18 AM	26838	Mouse Left Click	User left click in "Survey | Qualtrics Survey Software | Current Progress 42% - Google Chrome"
```

Figure C.7. Sample of tab delimited file parsed from XML in Figure C.6.

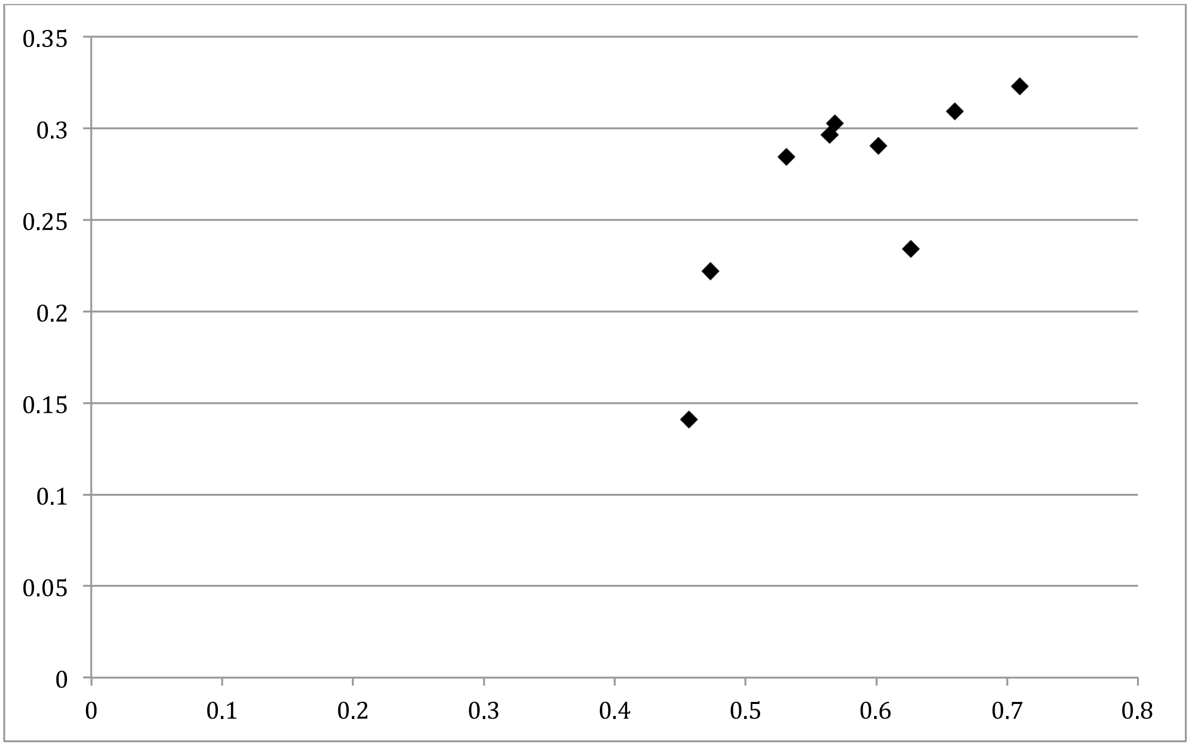


Figure C.8. Scatter plot of attribution vs. recognition by site. Attribution values are on y-axis, recognition values on x-axis.

References

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, *24*(4), 665–694.

Agarwal, R., & Venkatesh, V. (2002). Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability. *Information Systems Research*, *13*(2), 168–186.

Agnew, J. (2011). Space and place. In J. Agnew & D. N. Livingstone (Eds.), (pp. 316–330). London: Sage Publications.

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research*, *13*(4), 411–454.

Albert, T. C., Goes, P. B., & Gupta, A. (2004). GIST: A Model for Design and Management of Content and Interactivity of Customer-Centric Web Sites. *MIS Quarterly*, *28*(2), 161–182.

Allison, P. D. (2012). *Logistic Regression Using SAS: Theory and Application*. Cary, N.C.: SAS Institute.

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Digital, Inc.

Andrews, R. L., & Srinivasan, T. C. (1995). Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research*, *32*(1), 30–41.

Animesh, A., Ramachandran, V., & Viswanathan, S. (2010). Research Note -- Quality Uncertainty and the Performance of Online Sponsored Search Markets: An Empirical Investigation. *Information Systems Research*, *21*(1), 190–201.

Axelrod, J. N. (1968). Attitude measures that predict purchase. *Journal of Advertising Research*, *8*(1), 3–17.

Bagozzi, R. P., & Silk, A. C. (1983). Recall, Recognition, and the Measurement of Memory for Print Advertisements. *Marketing Science*, *2*(2), 95–134.

Baker, W., Hutchinson, J. W., & Moore, D. (1986). Brand Familiarity and Advertising: effects on the evoked set and brand preference. In R. J. Lutz (Ed.), (Vol. 13, pp. 637–642). Provo, UT: Association for Consumer Research.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bearden, W. O., Netemeyer, R. G., & Teel, J. E. (1989). Measurement of consumer susceptibility to interpersonal influence. *Journal of Consumer Research*, *15*(4), 473–481.

Belli, R. F., Lindsay, D. S., Gales, M. S., & McCarthy, T. T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Memory & Cognition*, *22*(1), 40–54.

Benbasat, I. (2006). Human-Computer Interaction for Electronic Commerce. In D. F. Galletta & P. Zhang (Eds.), . New York: M E Sharpe Inc.

Biehal, G., & Chakravarti, D. (1983). Information accessibility as a moderator of consumer choice. *Journal of Consumer Research*, *10*(1), 1–14.

Blackwell, R. D., Miniard, P. W., & Engel, J. F. (2006). *Consumer Behaviour*. Thomson.

Boudreau, M. C., Gefen, D., & Straub, D. W. (2001). Validation in Information Systems Research: A State-of-the-Art Assessment. *MIS Quarterly*, *25*(1), 1–16.

Brodie, R. J., Hollebeek, L. D., Juric, B., & Ilic, A. (2011). Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research. *Journal of Service Research*, *14*, 252–271.

Browne, G. J., Pitts, M. G., & Wetherbe, J. C. (2007). Cognitive Stopping Rules for Terminating Information Search in Online Tasks. *MIS Quarterly*, *31*(1), 89–104.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science*, *46*(4), 563–585.

Buhrmester, M., & Kwang, T. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

Buttimer, A. (1976). Grasping the dynamism of lifeworld. *Annals of the Association of American Geographers*, *66*, 277–292.

Byström, K., & Järvelin, K. (1995). Task Complexity Affects Information Seeking and Use. *Information Processing & Management*, *31*(2), 191–213.

Campbell, D. E., & Wells, J. D. (2013). Breaking the Ice in B2C Relationships: Understanding Pre-Adoption E-Commerce Attraction. *Information Systems Research*, *24*(2), 219–238.

Campbell, D. J. (1988). Task Complexity: A Review and Analysis. *Academy of Management Review*, *13*(1), 40–52.

Campbell, M. C., & Keller, K. L. (2003). Brand familiarity and advertising repetition effects. *Journal of Consumer Research*, *30*(2), 292–304.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Card, S. K., Pirolli, P., Van Der Wege, M., Morrison, J. B., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001). *Information scent as a driver of Web behavior graphs: results of a protocol analysis method for Web usability*. Presented at the Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, New York, USA.

Collins, C. J. (2007). The interactive effects of recruitment practices and product awareness on job seekers' employer knowledge and application behaviors. *Journal of Applied Psychology*, *92*, 180–190.

Coyle, J. R., & Thorson, E. (2001). The effects of progressive levels of interactivity and vividness in web marketing sites. *Journal of Advertising*, *30*(3), 65–77.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, N.Y.: Harper and Row.

Davenport, C. (2013, March 18). *EU privacy regulators take aim at Google privcy policy*. *Reuters.com*.

Davies, S. P. (2003). Initial and concurrent planning in solutions to well-structured problems. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*, 1147–1164.

Dennett, D. C. (1991). *Consciousness explained*. Boston, Mass.: Little, Brown & Co.

Dunn, B. K., Ramasubbu, N., Galletta, D. F., & Lowry, P. B. (2014). *Time and Territory: Optimal User Behavior for Brand Engagement from the Website Owner's Perspective*. Working Paper.

Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). Stanford University.

Econsultancy.com. (2011). *State of Search Marketing Report 2011* (p. 131).

Edelman, B., & Ostrovsky, M. (2007). Strategic bidder behavior in sponsored search auctions. *Decision Support Systems*, *43*(1), 192–198.

Edelman, B., Ostrovsky, M., & Schwarz, M. (2007). Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords. *The American Economic Review*, *97*(1), 242–259.

eMarketer. (2012). *US Digital Ad Spending to Top $37 Billion in 2012 as Market Consolidates*. *eMarketer*.

Fang, X., Hu, P. J. H., Chau, M., Hu, H. F., & Yang, Z. (2012). A Data-Driven Approach to Measure Web Site Navigability. *Journal of Management Information Systems*, *29*, 173–212.

Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: an empirical study. *Journal of the American Society for Information Science and Technology*, *52*(12), 1049–1066.

Ford, N., Miller, D., & Moss, N. (2005). Web search strategies and human individual differences: Cognitive and demographic factors, Internet attitudes, and approaches. *Journal of the American Society for Information Science and Technology*, *56*(7), 741–756.

Galletta, D. F., Henry, R. M., McCoy, S., & Polak, P. (2006). When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth. *Information Systems Research*, *17*(1), 20–37.

Gardner, M. P. (1985). Does attitude toward the ad affect brand attitude under a brand evaluation set? *Journal of Marketing Research*, *22*(2), 192–198.

Gartner. (2013). *Key Findings from U.S. Digital Marketing Spending Survey, 2013*. *Gartner*.

Gefen, D. (2002). Customer loyalty in e-commerce. *Journal of the Association for Information Systems*, *3*, 27–51.

Gefen, D., Karahanna, E., & Straub, D. W. (2003a). Inexperience and experience with online stores: the importance of TAM and trust. *IEEE Transactions on Engineering Management*, *50*(3), 307–321.

Gefen, D., Karahanna, E., & Straub, D. W. (2003b). Trust and TAM in Online Shopping: An Integrated Model. *MIS Quarterly*, *27*(1), 51–90.

Gensch, D. H. (1987). A Two-Stage Disaggregate Attribute Choice Model. *Marketing Science*, *6*(3), 223–239.

Ghose, A., & Yang, S. (2009). An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Management Science*, *55*(10), 1605–1622.

Gilbride, T. J., & Allenby, G. M. (2004). A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science*, *23*(3), 391–406.

Göker, A., & He, D. (2000). Analysing Web Search Logs to Determine Session Boundaries for User-Oriented Learning. In P. Brusilovsky, O. Stock, & C. Strapparava (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems* (Vol. 1892, pp. 319–322). Berlin: Springer Verlag.

Guo, C. (2001). A review on consumer external search: amount and determinants. *Journal of Business and Psychology*, *15*(3), 505–519.

Gustafson, P. (2001a). Meanings of place: Everyday experience and theoretical conceptualizations. *Journal of Environmental Psychology*, *21*(1), 5–16.

Gustafson, P. (2001b). Roots and Routes Exploring the Relationship between Place Attachment and Mobility. *Environment and Behavior*, *33*, 667–686.

Haley, R. I., & Case, P. B. (1979). Testing thirteen attitude scales for agreement and brand discrimination. *The Journal of Marketing*, *43*(4), 20–32.

Harrison, A. A. (1977). Mere Exposure. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*. New York, N.Y.: Academic Press.

Heinström, J. (2003). Five personality dimensions and their influence on information behaviour. *Information Research*, *9*(1), 1–24.

Hoffman, D. L., & Novak, T. P. (1996). Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. *Journal of Marketing*, *60*(3), 50–68.

Huang, M.-H. (2003). Designing website attributes to induce experiential encounters. *Computers in Human Behavior*, *19*, 425–442.

Hui, K.-L., Teo, H. H., & Lee, S.-Y. T. (2007). The value of privacy assurance: an exploratory field experiment. *MIS Quarterly*, *31*(3), 19–33.

ITU. (2014). *2005-2014 ICT data for the world, by geographic regions and by level of development*. International Telecommunications Union. Retrieved from http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2014/ITU_Key_2005-2014_ICT_data.xls

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management*, *44*(3), 1251–1266.

Jansen, B. J., & Mullen, T. (2008). Sponsored search: an overview of the concept, history, and technology. *International Journal of Electronic Business*, *6*(2), 114–131.

Jiang, Z., & Benbasat, I. (2007). Research Note—Investigating the Influence of the Functional Mechanisms of Online Product Presentations. *Information Systems Research*, *18*(4), 454–470.

Jiang, Z., & Benbasat, I. (2007). The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding. *MIS Quarterly*, *31*(3), 475–500.

Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). On the depth and dynamics of online search behavior. *Management Science*, *50*(3), 299–308.

Johnson, E. J., & Russo, J. E. (1981). Product familiarity and learning new information. *Advances in Consumer Research*, *8*(1), 151–155.

Katz, M. A., & Byrne, M. D. (2003). Effects of scent and breadth on use of site-specific search on e-commerce Web sites. *ACM Transactions on Computer-Human Interaction*, *10*(3), 198–220.

Keller, K. L. (2008). *Strategic Brand Management: Building, Measuring, and Managing Brand Equity*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Kent, R. J., & Allen, C. T. (1994). Competitive interference effects in consumer memory for advertising: the role of brand familiarity. *The Journal of Marketing*, *58*(3), 97–105.

Kent, W. (1978). *Data and Reality*. New York: North-Holland Publishing Company.

Kessler, G. (2013). *How much did HealthCare.gov cost? (Part 2)*. *Washington Post*.

Kim, T., & Biocca, F. (1997). Telepresence via Television: Two Dimensions of Telepresence May Have Different Connections to Memory and Persuasion. *Computer-Mediated Communication*, *3*.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, *9*(2), 137–163.

Kohli, R., Devaraj, S., & Mahmood, M. A. (2004). Understanding Determinants of Online Consumer Satisfaction: A Decision Process Perspective. *Journal of Management Information Systems*, *21*(1), 115–136.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a Psychology of Memory Accuracy. *Annual Review of Psychology*, *51*, 481–537.

Kotler, P. (1997). *Marketing Management: Analysis, Planning, Implementation, and Control*. Englewood Cliffs, N.J.: Prentice Hall.

Koufaris, M. (2002). Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. *Information Systems Research*, *13*(2), 205–223.

Krug, S. (2009). *Don't Make Me Think*. Pearson Education.

Laroche, M., Kim, C., & Zhou, L. (1996). Brand familiarity and confidence as determinants of purchase intention: an empirical test in a multiple brand context. *Journal of Business Research*, *37*, 115–120.

Lavidge, R. J., & Steiner, G. A. (1961). A Model for Predictive Measurements of Advertising Effectiveness. *Journal of Marketing*, *25*(6), 59–62.

Lederer, A. L., Maupin, D. J., Sena, M. P., & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems*, *29*(3), 269–282.

Lee, A. Y., Keller, P. A., & Sternthal, B. (2010). Value from Regulatory Construal Fit: The Persuasive Impact of Fit between Consumer Goals and Message Concreteness. *Journal of Consumer Research*, *36*(5), 735–747.

Lee, M. K. O., & Turban, E. (2001). A Trust Model for Consumer Internet Shopping. *International Journal of Electronic Commerce*, *6*(1), 75–91.

Lewis, E. S. E. (1903). Advertising Department: Catch-Line and Argument. *The Book-Keeper*, *15*, 124–128.

Lewis, P. (1979). Defining Sense of Place. In W. P. Prenshaw & J. O. McKee (Eds.), (pp. 24–46). Jackson, Miss.: University of Mississippi.

Li, D., Browne, G. J., & Wetherbe, J. C. (2006). Why Do Internet Users Stick with a Specific Web Site? *International Journal of Electronic Commerce*, *10*, 105–141.

Liu, J., Zhang, S., & Yang, J. (2004). Characterizing web usage regularities with information foraging agents. *IEEE Transactions on Knowledge and Data Engineering*, *16*, 566–584.

Liu, Y. (2003). Developing a Scale to Measure the Interactivity of Websites. *Journal of Advertising Research*, *43*, 207–216.

Liu, Y., & Shrum, L. J. (2002). What is Interactivity and is it Always Such a Good Thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness. *Journal of Advertising*, *31*(4), 53–64.

Lowry, P. B., Cao, J., & Everard, A. (2011). Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: The case of instant messaging in two cultures. *Journal of Management Information Systems*, *27*(4), 163–200.

Lowry, P. B., Gaskin, J., Twyman, N., Hammer, B., & Roberts, T. L. (2013). Proposing the hedonic-motivation system adoption model (HMSAM) to increase understanding of adoption of hedonically motivated systems. *Journal of the Association for Information Systems*.

Lowry, P. B., Romano, N. C., Jenkins, J. L., & Guthrie, R. (2009). The CMC Interactivity Model: How Interactivity Enhances Communication Quality and Process Satisfaction in Lean-Media Groups. *Journal of Management Information Systems*, *26*(1), 155–195.

Lowry, P. B., Vance, A., Moody, G., & Beckman, B. (2008). Explaining and predicting the impact of branding alliances and web site quality on initial consumer trust of e-commerce web sites. *Journal of Management Information Systems*, *24*(4), 199–224.

Lunden, I. (2013, September 30). Digital Ads Will Be 22% of All U.S. Ad Spend in 2013, Mobile Ads 3.7%; Total Global Ad Spend in 2013 $503B. *TechCrunch*. Retrieved from http://techcrunch.com/2013/09/30/digital-ads-will-be-22-of-all-u-s-ad-spend-in-2013-mobile-ads-3-7-total-gobal-ad-spend-in-2013-503b-says-zenithoptimedia/

Machleit, K. A., & Wilson, R. D. (1988). Emotional feelings and attitude toward the advertisement: The roles of brand familiarity and repetition. *Journal of Advertising*, *17*, 27–35.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*.

Mathwick, C., & Rigdon, E. (2004). Play, Flow, and the Online Search Experience. *Journal of Consumer Research*, *31*(2), 324–332.

McCoy, S., Everard, A., Polak, P., & Galletta, D. F. (2007). The effects of online advertising. *Communications of the ACM*, *50*, 84–88.

McKinney, V., Yoon, K., & Zahedi, Fatemeh Miriam. (2002). The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. *Information Systems Research*, *13*(3), 296–315.

McKnight, D. H., & Choudhury, V. (2006). Distrust and trust in B2C e-commerce: Do they differ? *Proceedings of the 8th International Conference on Electronic Commerce*, 482–491.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

Menard, S. (2002). *Applied Logistic Regression*. Thousand Oaks, Calif.: Sage.

Mennecke, B. E., Triplett, J. L., & Hassall, L. M. (2011). An examination of a theory of embodied social presence in virtual worlds. *Decision Sciences*, *42*(2), 413–450.

Milgrom, P. (2010). Simplified mechanisms with an application to sponsored-search auctions. *Games and Economic Behavior*, *70*(1), 62–70.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Miller, G. A. (1983). Informavores. In F. Machlup & U. Mansfield (Eds.), (pp. 111–113). New York: Wiley.

Mithas, S., Ramasubbu, N., Krishnan, M. S., & Fornel, C. (2007). Designing web sites for customer loyalty across business domains: a multilevel analysis. *Journal of Management Information Systems*, *23*, 97–127.

Mollen, A., & Wilson, H. (2010). Engagement, telepresence and interactivity in online consumer experience: Reconciling scholastic and managerial perspectives. *Journal of Business Research*, *63*, 919–925.

Montgomery, A. L., Li, S., & Srinivasan, K. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, *23*, 579–595.

Moorthy, S., Ratchford, B. T., & Talukdar, D. (1997). Consumer information search revisited: Theory and empirical analysis. *Journal of Consumer Research*, *23*(4), 263–277.

Morison, B. (2002). *On Location*. Oxford: Clarendon Press.

Nadkarni, S., & Gupta, R. (2007). A task-based model of perceived website complexity. *MIS Quarterly*, *31*, 501–524.

Nielsen, J. (2000). *Designing web usability*. Peachpit Press.

Novak, T. P., Hoffman, D. L., & Duhachek, A. (2003). The Influence of Goal-Directed and Experiential Activities on Online Flow Experiences. *Journal of Consumer Psychology*, *13*, 3–16.

Novak, T. P., Hoffman, D. L., & Yung, Y.-F. (2000). Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Marketing Science*, *19*(1), 22–42.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill Humanities/Social Sciences/Languages.

O'Brien, T. (1971). Stages of consume decision making. *Journal of Marketing Research*, *8*(3), 283–289.

Palmer, J. W. (2002). Web Site Usability, Design, and Performance Metrics. *Information Systems Research*, *13*(2), 151–167.

Pavlou, P. A. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, *7*(3), 101–134.

Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: an extension of the theory of planned behavior. *MIS Quarterly*, *30*, 115–143.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford: Oxford University Press.

Pechman, C., & Stewart, D. W. (1988). Advertising Repetition: A Critical Review of Wearin and Wearout. *Current Issues and Research in Advertising*, *11*(1-2), 285–329.

Pemberton, S. (2003). Hotel heartbreak. *Interactions*, *10*, 64.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford: Oxford University Press.

Pirolli, P., & Card, S. K. (1999). Information foraging. *Psychological Review*, *106*, 643–675.

Punj, G. N., & Staelin, R. (1983). A model of consumer information search behavior for new automobiles. *Journal of Consumer Research*, *9*(4), 366–380.

Rafaeli, S. (1990). Interacting with Media: Para-Social Interaction and Real Interaction. In B. D. Ruben & L. A. Lievrouw (Eds.), (Vol. 3). New Brunswick, NJ: Transaction Publishers.

Relph, E. (1976). *Place and Placelessness*. London: Pion.

Roediger, H. L. I. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, *8*(3), 231–246.

Rogers, E. M. (1986). *Communication Technology*. New York: The Free Press.

Ross, N., & Wolfram, D. (2000). End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine. *Journal of the American Society for Information Science*.

Saunders, C., Rutkowski, A. F., von Genuchten, M., & Vogel, D. (2011). Virtual space and place: theory and test. *MIS Quarterly*, *35*, 1079–1098.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182–203.

Schacter, D. L., & Dodson, C. S. (2001). Misattribution, false recognition and the sins of memory. *Philosophical Transactions of the Royal Society*, *356*(1413), 1385–1393.

Shih, C. F. (1998). Conceptualizing consumer experiences in cyberspace. *European Journal of Marketing*, *32*, 655–663.

Shrivastava, P. (1987). Rigor and practical usefulness of research in strategic management. *Strategic Management Journal*, *8*(1), 77–92.

Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*, 99.

Simon, H. A. (1960). *The New Science of Management Decision* (Vol. 3). New York, N.Y.: Harper & Brothers.

Simon, H. A. (1971). Designing organizations in an information-rich world. In M. Greenberger (Ed.), (pp. 37–53). Baltimore, Md.: Johns Hopkins University Press.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, *4*, 181–201.

Simon, H. A. (1981). *The sciences of the artificial*. Cambridge: The Massachusetts Institute of Technology.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.

Song, J. H., & Zinkhan, G. M. (2008). Determinants of Perceived Web Site Interactivity. *Journal of Marketing*, *72*, 99–113.

Spink, A., Wolfram, D., & Jansen, B. J. (2001). Searching the web: The public and their queries. *Journal of the Association for Information Science and Technology*, *52*(3), 226–234.

Srinivasan, Narasimhan, & Ratchford, Brian T. (1991). An Empirical Test of a Model of External Search for Automobiles. *Journal of Consumer Research*, *18*, 233–242.

Steelman, Z., Hammer, B. I., & Limayem, M. (2014). Data Collection in the Digital Age: Innovative Alternatives to Student Samples. *MIS Quarterly*, *Forthcoming*.

Steuer, J. (2006). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*, 73–93.

Stewart, D. W. (1992). Speculations on the future of advertising research. *Journal of Advertising*, *21*, 1–18.

Stigler, G. J. (1961). The economics of information. *The Journal of Political Economy*, *69*, 213–225.

Tanzer, M. (2014, May 15). *Exclusive: New York Times Internal Report Painted Dire Digital Picture*. *BuzzFeed*. Retrieved May 23, 2014, from http://www.buzzfeed.com/mylestanzer/exclusive-times-internal-report-painted-dire-digital-picture?_ga=1.124475914.1811747779.1394659920

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology*, *83*, 268–277.

Thompson, D. (2014, May 15). *What the Death of Homepages Means for the Future of News*. *The Atlantic*. Retrieved May 23, 2014, from http://www.theatlantic.com/business/archive/2014/05/what-the-death-the-homepage-means-for-news/370997/

Trevino, L. K., & Webster, J. (1992). Flow in Computer-Mediated Communication Electronic Mail and Voice Mail Evaluation and Impacts. *Communication Research*, *19*, 539–573.

Tuan, Y. F. (1974). Space and Place: Humanistic Perspective. In S. Gale & G. Olsson (Eds.), . London: D. Reidel Publishing Company.

Tuan, Y. F. (1975). Place: an experiential perspective. *Geographical Review*, *65*, 151–165.

Tuan, Y. F. (1977). *Space and Place*. Minneapolis: University of Minnesota Press.

Tuan, Y. F. (2001). *Space and place: The perspective of experience*. Minneapolis: University of Minnesota Press.

Venkatesh, V., & Ramesh, V. (2006). Web and wireless site usability: understanding differences and modeling use. *MIS Quarterly*, *30*(1), 181–206.

Webster, J., & Ahuja, J. S. (2006). Enhancing the design of web navigation systems: The influence of user disorientation on engagement and performance. *MIS Quarterly*, *30*, 661–678.

Yao, S., & Mela, C. (2011). A Dynamic Model of Sponsored Search Advertising. *Marketing Science*, *30*(3), 447–468.

Zaichkowsky, J. L. (1994). The personal involvement inventory: Reduction, revision, and application to advertising. *Journal of Advertising*, *23*, 59–70.

Zajonc, R. B. (1968). Attitudinal Effects of Mere Exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–368.