Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth

Dennis F. Galletta, Raymond M. Henry, Scott McCoy, Peter Polak,

Information Systems Research
Vol. 17, No. 1, March 2006, pp. 20–37
ISSN 1047-7047 | EISSN 1526-5536 | 06 | 1701 | 0020
DOI 10.1287/isre.1050.0073

When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth

Dennis F. Galletta
Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, galletta@katz.pitt.edu

Raymond M. Henry
Department of Management, Clemson University, Clemson, South Carolina 29634, rhenry@clemson.edu

Scott McCoy
School of Business, College of William & Mary, Williamsburg, Virginia 23187-8795, scott.mccoy@mason.wm.edu

Peter Polak
School of Business Administration, University of Miami, Coral Gables, Florida 33124-6540, ppolak@miami.edu

Although its popularity is widespread, the Web is well known for one particular drawback: its frequent delay when moving from one page to another. This experimental study examined whether delay and two other website design variables (site breadth and content familiarity) have interaction effects on user performance, attitudes, and behavioral intentions. The three experimental factors (delay, familiarity, and breadth) collectively impact the cognitive costs and penalties that users incur when making choices in their search for target information. An experiment was conducted with 160 undergraduate business majors in a completely counterbalanced, fully factorial design that exposed them to two websites and asked them to browse the sites for nine pieces of information. Results showed that all three factors have strong direct impacts on performance and user attitudes, in turn affecting behavioral intentions to return to the site, as might be expected. A significant three-way interaction was found between all three factors indicating that these factors not only individually impact a user's experiences with a website, but also act in combination to either increase or decrease the costs a user incurs. Two separate analyses support an assertion that attitudes mediate the relationship of the three factors on behavioral intentions. The implications of these results for both researchers and practitioners are discussed. Additional research is needed to discover other factors that mitigate or accentuate the effects of delay, other effects of delay, and under what amounts of delay these effects occur.

Key words: electronic commerce; response time; website design; attitudes; performance
History: Cynthia Beath, Senior Editor; Iris Vessey, Associate Editor. This paper was with the authors 18 months for 3 revisions.

In spite of large investments in business-to-consumer electronic commerce, sales of online stores lag far behind those of brick-and-mortar stores. Online consumer spending represents only 2.2% of the $916.8 billion in total retail sales for the most recent quarter in the United States (Commerce Department 2005). At least part of the gap might be explained by a major difficulty users encounter when browsing through online stores—the pervasive delay that accompanies each attempt to view information. Delay is considered "the most commonly experienced problem with the Web" (Pitkow et al. 1998), leading many to call the Web the "worldwide wait" (Nah and Kim 2000).

Slow sites suffer sharp usage declines (Nielsen 1999a, Wonnacott 2000), and Zona Research Study (1999) estimates losses associated with download times of eight seconds and higher to be $4.35 billion annually. Building on research on system delay in many settings over the years, this study is aimed at developing a more complete understanding of delay, and investigates how related factors can mitigate the negative effects of delay on attitudes, behavioral intentions, and performance of users.

No "Silver Bullet"

One potential "silver bullet" is increasing the speed of access. High-speed access is indeed a seductive attempt at a cure, and U.S. users who are connected at high speed are now in the majority. However, broadband users still sometimes encounter significant delay

(Nielsen 1999b, Selvidge 2003), and page size and complexity continue to grow. Traffic growth continues to outpace network bandwidth upgrades, painting a gloomy picture for the future (Sears and Jacko 2000).

Many causes of delay are independent of the user's connection speed (Nielsen 1997, Rose et al. 1999, Nah and Kim 2000). The ultimate speed of loading a page is determined by the weakest link in the transmission path from the server, to the Internet backbone, to the Internet provider, and finally, to the user's desktop. Even without congestion, users can still find themselves in a global waiting line at a popular site, or the victim of server configuration errors that can even triple page-loading time (Koblentz 2000). In spite of broadband, "delays will be a concern for many years to come" (Sears and Jacko 2000, p. 49).

A second potential silver bullet is the use of a search engine to provide direct access to a single target page. Unfortunately, search engines provide many difficulties to users who try to exploit them. Keywords, terminology, and context make searching difficult (Olston and Chi 2003).[1] Search engines are difficult for many users (Gauch and Smith 1991, Lee 1999, Muramatsu and Pratt 2001), and the unique syntax of each exacerbates the situation (Lohse and Spiller 1998).

Although the obvious silver bullets seem unlikely to defeat delay, if users can proceed to their desired page without becoming lost in the site's structure or having to work through several layers, perhaps delay can be more tolerable or less noticeable. This paper focuses on how familiarity and breadth, which will be defined below, might dampen the ill effects of delay.

Delay, Familiarity, and Breadth

Familiarity and site breadth are hypothesized in this paper to interact with delay and dampen delay's negative impacts on users. The interaction is more interesting than each main effect considered separately. Substantial literature exists on each separate effect, but not on how they work together.

Familiarity with the terminology used in hyperlinks and with the structure of a site can reduce a user's trial-and-error behavior (Edwards and Hardman 1989). Users who are familiar with the terminology and structure are less likely to become confused, or "lost," and can minimize the number of clicks—and overall delay—to reach the target. Breadth of a site is another favorable design goal. Site designs with many links per page and few levels have been found to be preferable to deep sites with fewer links per page. The rule appears to be that no more than two levels of traversal should be required (Zaphiris and Mtei 1997, Larson and Czerwinski 1998). A broad site requires only a few clicks to reach the target, also minimizing the delays that accompany the required traversal from top to bottom.

Although this study focuses on dampening of delay's effects, one might also approach the problem from the "other side." One might posit an exacerbation effect of undesirable site attributes, in which unfamiliarity and depth exacerbate the effects of delay. Each point of view is legitimate, but each assumes a different starting point. If one assumes that most sites are unfamiliar and deep, then the interaction is one of dampening the effects of delay. Conversely, if one assumes that most sites are appropriately familiar and broad, then the proper effect to study is exacerbation by introducing *un*familiarity and *depth*. In this study, we take a somewhat pessimistic stance and assume that most sites could stand significant improvements in familiarity and breadth to dampen the effects of delay. The factors lead to the following research question:

How do delay, familiarity, and site breadth interact to influence attitudes, performance, and behavioral intentions in online product information searches?

This question should be of interest to researchers who wish to understand how these factors interact to form a predictive model. It should also be of interest to designers who must make choices in site design, such as evaluating the trade-offs between cluttering a screen with many links per page in a "broad" site, or providing a "clean" look and forcing users to drill

[1] Olston and Chi (2003) provide more complete explanations of the difficulties posed by search engines. Often, users do not know a priori precisely what options are available before browsing through the site. Some terms are not expressed precisely enough for use in a search engine (for example, it is difficult to search for a printer that operates at least 20 pages per minute if it is expressed variously as "21 ppm," "22 pages/minute," or "25 pgs/min"). Finally, information and context about the path is missing if a user obtains direct access to the target page. In their words, "providing site search should not be used to compensate for poor menu design" (p. 198).

down very deep. Designers also must choose between taking up space and explaining technical terminology or providing a less cluttered interface that assumes some level of user knowledge.

Theoretical Foundations

Individuals must navigate within the structure of a site to find information (Chi et al. 2001, Bernard 2002). Structure is primarily the arrangement of links between the various pages, or nodes, that a user must traverse to reach a desired destination (Chi et al. 2000). Users are faced with a decision point at each node of a site, and they evaluate the likely effort and probability of reaching their target via each link (candidate path). Previous research suggests that individuals develop strategies to limit their required cognitive effort (Johnson and Payne 1985, Todd and Benbasat 1999), seeking "low-cost" strategies. The sometimes conscious, sometimes unconscious assessment of cost is therefore an important part of users' experiences with websites. Sites providing lower cost—in terms of total time and effort—will be preferred over those with higher cost.

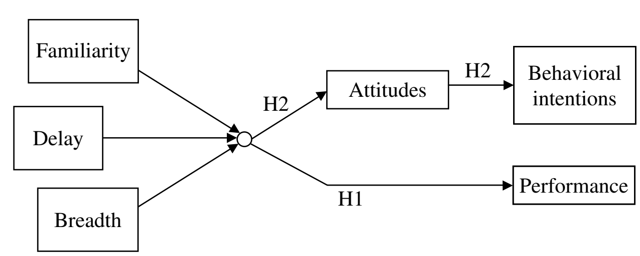
An explanation for this preference is provided by Pirolli and Card's (1999) Information Foraging Theory (Larson and Czerwinski 1998, Katz and Byrne 2003), which asserts that individuals employ strategies to maximize their hit rate and minimize the costs of finding information they seek (Katz and Byrne 2003). A key component of this theory is the concept of information scent, defined as "the 'imperfect' perception of the value, cost, or access path of information sources obtained from proximal cues" (Pirolli and Card 1995, p. 646) such as links.

These proximal cues allow a user to estimate the location of the "distal" content (on further pages) and determine an appropriate path (Chi et al. 2000, p. 2). When the scent is strong, users can easily find target information, expending little cognitive effort and avoiding wrong paths. According to Card et al. (2001), the scent is provided by words used in links, images, page layout, or other contextual cues. All of these elements are part of a site's structure (Korthauer and Koubek 1994), suggesting a relationship between a site's structure, the amount of information scent that is provided, and the overall cost of navigating to the right location.

If the scent is too low, users can become disoriented or "lost" in the structure. Otter and Johnson (2000) listed possible reasons for becoming lost, asking respondents to mark those they had experienced. Two-thirds of the responses covered five items in decreasing frequency: site size, unfamiliar material, predictability of link destinations, unfamiliar terminology, and hierarchical structure. Two important dimensions seemed to emerge: site depth (addressed by the first and last of the five) and lack of familiarity (addressed by the others). In more open-ended questions, when asked to list reasons for becoming lost, responses again provided support for the critical nature of breadth and familiarity in preventing disorientation. Both dimensions affect the scent and, thus, the ability to find target information. As highlighted above, there is a relationship between a site's structure and familiarity and the amount of information scent that is provided.

Delay directly raises the cost of navigation and further reduces the scent by pushing the target further into the future. Becoming disoriented while traversing many levels of a deep site and making incorrect choices when sites are unfamiliar raises the cost of browsing significantly by requiring backtracking and seeking a path with a stronger scent. On the other hand, when delay is a significant problem, navigation through a site that is broad and familiar can reduce the negative impacts by lowering the total costs from delay that are imposed on the user. As a result, we believe that focusing only on any one of these factors presents an incomplete picture of the costs of browsing through a site toward a definite target. We focus on the interaction of all three factors to predict user attitudes, performance, and behavioral intentions. That is, we provide a study that investigates how familiarity and breadth dampen the effects

Figure 1 Research Model



of delay by increasing the scent of the target page. The research model is presented in Figure 1. In the following sections, the elements of the model are described and hypotheses about the relationships among these elements are developed.

Delay. Response time has been shown consistently to be a major factor in overall usability, both in studies of the Web (Nielsen 1999b; Rose et al. 1999, 2001; Rose and Straub 2001; McKinney et al. 2002; Torkzadeh and Dillon 2002) as well as large-scale time-sharing systems (Kuhmann 1989, Shneiderman 1998).

Effects of Delay. Prior research has shown an inverse relationship between response time and user performance (Barber and Lucas 1983, Butler 1983) and productivity (Dannenbring 1983, Martin and Corl 1986). Long delays lead to frustration (Doherty and Kelisky 1979), dissatisfaction (Lee and MacGregor 1985), feelings of being lost (Sears et al. 2000), and giving up (Nah 2002). Negative impacts are strongest when delays are longer than expected, occur in unpredictable patterns, or are unaccompanied by duration or status information (Dellaert and Kahn 1999, Polak 2002).

Tolerable Level of Delay. Early studies suggested that the ideal response time is two seconds (Miller 1968), although such short response times might be impossible for developers to achieve, given the abundant and varied sources for delay.

Estimates of maximum tolerable delay and specific effects of that delay have shown some variation in previous studies, but overall, those effects are always negative. As delays exceed six seconds, performance, attitudes, and behavioral intentions seem to "bottom out" (Galletta et al. 2004). At eight seconds, users suffer psychological and performance consequences (Kuhmann 1989). At 10 seconds, they lose interest (Ramsay et al. 1998). At 12 seconds, users lose patience (Hoxmeier and DiCesare 2000). At 15 seconds, delay becomes disruptive (Shneiderman 1998). At 38 seconds, users will abandon a task (Nah 2002). Together, these results suggest that as delays increase, users become more impatient and display this impatience in different ways.

It is striking that such relatively short delays in prior research have been shown to produce significant results in both user attitudes and performance. Web users often must deal with much longer wait times (Ramsay et al. 1998). From a cognitive perspective, increased delays increase the cognitive effort that users must expend. Delays directly increase the cost that a user incurs for each choice that is made in finding target information.

Breadth. Website designs can vary in the number of selections that must be made on the way to a target node. Given a fixed number of end nodes in a website, a designer can elect to use a "broad" strategy, increasing the number of hyperlinks on each page while decreasing the number of clicks and page loads, or a "deep" strategy with fewer links per page but more hierarchical levels. The trade-off between breadth and depth has been widely studied in menu design for information acquisition (for example, see Miller 1981, Kiger 1984, Landauer and Nachbar 1985, Norman and Chin 1988, Parkinson et al. 1988) and hypermedia (Edwards and Hardman 1989), but only recently in Web design (Zaphiris and Mtei 1997, Chae and Kim 2003, Katz and Byrne 2003).

Much of this work has focused on determining the optimal number of items per level (Larson and Czerwinski 1998). Although some variation has occurred in research findings, most studies listed above have determined that breadth is preferred to depth (Jacko et al. 1995).

The theory of task complexity (Frese 1987) can explain this preference (Jacko and Salvendy 1996): A deeper structure increases the number of decisions that must be made, which in turn increases complexity, response time, and the number of errors. Additionally, the information scent, the certainty that the chosen path is the correct one, decreases as depth increases (Bernard 2002), perhaps because of a weaker association between the descriptor and the target (Larson and Czerwinski 1998). These explanations are consistent with Shneiderman's (1998) observation: "when the depth goes to four or five (levels), there is a good chance of users becoming lost or disoriented" (p. 249).

Although increasing the breadth of a site requires fewer clicks by the user to reach the desired page, providing too many links on a page can result in clutter, making it difficult to read or comprehend (Larson and Czerwinski 1998). Most laboratory studies consider eight to nine items per menu to be a "broad" structure and two to three items per menu to be a "deep" structure, but even when those numbers are varied, the preference for breadth over depth remains (Katz and Byrne 2003).

To summarize, an ideal depth or breadth of a website cannot be determined, although most studies have found that users prefer—and perform best with—a "broad" structure with eight to nine items per menu in interactive systems. These choices can have direct implications on the complexity of a site, the information scent that is provided, and the number of decisions that must be made. The choices impact not only a user's ability to find the target, but also the search cost incurred by the user.

Familiarity. Familiarity with website content has been shown to be an important determinant of shopping effectiveness, but familiarity is difficult to control because of widespread heterogeneity of users (Chau et al. 2000). We therefore define familiarity in terms of users' *understanding of the terminology presented on the site*. Familiarity allows them to discern each move toward reaching their goal. Using meaningful categories has been described as important for website ease of use (Lederer et al. 2000) and critical for effective e-commerce (Turban and Gehrke 2000). These structural elements directly impact information acquisition costs, cognitive effort required, and information scent.

Familiarity with the terms used to define the paths on a site should increase the information scent provided by the paths. Pages higher on the hierarchy often depict broad identifiers for which the user must make inclusion judgments to reach the appropriate page at lower levels (Paap and Roske-Hofstrand 1988, Pardue and Landry 2001, Katz and Byrne 2003). If users do not understand the category names and partitions used by designers, they will take longer to select a path (Somberg and Picardi 1983) and make more mistakes in reaching their goal (Dumais and Landauer 1983). The higher scent of a familiar path will allow users to discover that they are in an incorrect branch, and will enable an efficient strategy to head up and over toward the desired one. Delays caused by searching for the correct path on what can become a trial-and-error basis (Schwartz and Norman 1986) will be minimized when category names are familiar (Paap and Roske-Hofstrand 1988).

Familiarity will minimize the disorientation that occurs when users become lost. Such disorientation is so widespread and serious that Edwards and Hardman (1989) call it "hypertext's major limiting factor" (p. 62). A cognitive explanation of this is that users spend some of their limited processing capabilities on navigation that could otherwise be spent on attending to comprehension of the content (Thuring et al. 1995). This cognitive overhead should obviously be minimized or the site will become unusable.

Quite often, firms organize websites according to their own internally generated product-line categories. For example, one very large consumer electronics firm once included three paths to its camcorders. The links were "Camcorders," "Handycams," and "Professional Camcorders." All were video camcorders, but in different price, size, and/or feature groupings. A customer might be looking for a particular model or format, but familiarity with the groupings would allow quick access, preventing the need for a "brute force" search through all of the model families.

Dampening Effects of Breadth and Familiarity on Delay. In narrative form, the model asserts that breadth and familiarity dampen the effects of delay on two important outcomes: performance and intentions to return. The effect on behavioral intentions, however, is expected to be mediated by attitudes. The hypothesized interactions are based on the extremely high cost of foraging under the most unfavorable combination of conditions. Although each factor discussed above is important, our ultimate intention is to posit and test the three-way interaction to demonstrate the dampening effect. The conditions of breadth and familiarity are expected to combine to alleviate the cost (and attendant negative consequences) of delay. The argument will be constructed by examining two factors first, then adding the third. Statistically speaking, the three-way test will supplant the main effects and lower-order interactions.

Breadth and Familiarity

While decreasing breadth and increasing depth will decrease the rate at which information is acquired (Larson and Czerwinski 1998, Pirolli and Card 1999), decreasing familiarity with the terminology on the site will introduce trial-and-error foraging at the same time. A familiar site with L levels to traverse (from the main index page to a bottom-level page) would require an expected average of C clicks to reach the bottom node as follows:

$$C_{fam} = L.$$

By definition, a perfectly familiar site supports error-free navigation; therefore, the number of clicks to reach a bottom-level page is equal to the number of levels the user must traverse, with no backtracking. For example, if there are five levels from top to bottom, a user who starts at Level 1 needs to traverse four levels to reach Level 5, the goal.

A perfectly unfamiliar site, however, without learning, would require a "brute force" systematic (next neighbor) search with backtracking. Such a search would require many more clicks to reach the correct bottom node. The expected number of clicks to reach the desired node is a function of the number of links per page and the number of levels

$$C_{unfam} = \left(\sum_{i=1}^{L-1} n^i\right) + 1$$

where n = the number of links per page (assumed to be constant throughout) and L = the number of levels to traverse.

Counting from the level under the home page, at each level i, there are n^i intermediate pages. In a brute force search, the user would traverse each of these pages, hoping to find a direct link to the bottom-level goal. Each unsuccessful search would normally stop at level $L-1$—that is, one level above the bottom—in the absence of the desired link. In other words, if a user is looking for a certain model camcorder, and arrives at a page including only links to the pages of other models, he or she would not need to load each page containing the incorrect model, and would move on to another promising page. Ultimately, when a link to the correct bottom page is found, he or she would need to click one last time to get there, hence the +1 at the end of the formula.

Traversing through the structure requires backtracking after each "dead end," suggesting two clicks per intermediate page (one down and one back up to take the next branch). However, on average, a randomly placed goal would be found halfway through the search. Therefore, the effect of backtracking is cancelled by finding the goal halfway through the task (multiplying by two clicks per page, then dividing by two), so the formula provides the average number of clicks to arrive at the goal.

Table 1 Average Number of Clicks Predicted for Navigating to a Given Node

	Deep (4 levels to traverse with 3 links each)	Broad (2 levels to traverse with 9 links each)
Familiar site	4	2
Unfamiliar site	40	10

If a site has 81 bottom-level pages, two options might be to arrange them into 4 levels each with 3 links ($3^4 = 81$) or into 2 levels each with 9 links ($9^2 = 81$). The differences between the expected number of clicks in a familiar and unfamiliar site using each option are striking (see Table 1). Depth increases the number of clicks by a factor of 2 to 4, while unfamiliarity increases it by a factor of 5 to 10. Each is a significant difference.

Adding the Delay Factor

The most striking conclusion of Table 2, which shows the predicted amount of time when navigating different types of sites, is the combined effect of breadth and familiarity to reduce the number of clicks to reach the goal, thus lessening by more than an order of magnitude the impact of delay. As the delay with each click increases (within each table cell), the differences among the cells increase concomitantly. If pages loaded instantaneously, the expected total delay (for all clicks to the goal) would be zero in all cells. However, if a site incurred an 8-second delay, a brute-force search in the broad site would impose a total delay of 80 seconds (1.3 minutes). In the deep site, the expected total delay would leap to 320 seconds (5.3 minutes). In a familiar site, the user would only

Table 2 Delay Predicted for Navigating to a Given Node, Zero- or Eight-Second Delay

	Deep (4 levels to traverse with 3 links each) 0-second delay/8-second delay	Broad (2 levels to traverse with 9 links each) 0-second delay/8-second delay
Familiar site	0 seconds/32 seconds	0 seconds/16 seconds
Unfamiliar site	0 seconds/320 seconds	0 seconds/80 seconds

need to click once at each level, with a total delay of 32 seconds and 16 seconds in the deep and broad site, respectively. Although previous researchers stopped short of examining these dimensions together, the mathematical and logical justification above suggests these factors interact with each other to increase the total costs imposed on users when searching for information.

Individual click delays impose a "micro" cost that is relatively fixed, but breadth and familiarity will reduce the number of required clicks, and reduce the *total* cost incurred by the user. Lower costs will make navigation within a website easier and can be expected to improve the performance of finding target information.

Hypothesis 1. *Breadth and familiarity will dampen the total cost of delay on overall performance of a user in a set of search tasks.*

A key Web design goal is to influence potential users to return later. Recent research suggests that the use of behavioral intentions is appropriate in a Web context as a proxy for design success (Song and Zahedi 2001). Total delay costs forced on users would be expected to impact intentions to return to a site. However, the formation of behavioral intentions is complex, and these effects may not be direct. Fishbein and Ajzen (1975) theorize attitudes to be a central antecedent of behavioral intentions, and the relationship between attitudes and behavior has seen significant attention (Ajzen 2001). Technology acceptance research (Davis 1989, Taylor and Todd 1995, Venkatesh et al. 2003) follows from this stream. Although the central TAM researchers (Venkatesh et al. 2003) later removed attitudes in favor of ease of use and usefulness as perhaps more detailed surrogates of attitudes, this study does not examine TAM factors and those antecedents of intentions are not used. Because of the importance of attitudes and the effects of very brief encounters in forming lasting attitudes about sites (Chiravuri and Peracchio 2003), the reduction of total delay facilitated by breadth and familiarity in site design is expected to improve users' attitudes toward a site, which would in turn increase their intentions to return.

Hypothesis 2. *Familiarity and breadth will dampen the negative effect of delay on behavioral intentions, but this relationship will be mediated by attitudes.*

Summary. To summarize, behavioral intentions and performance are identified as important outcomes of website usage. In this study, we focus on the ability of site breath and familiar content to dampen the negative impacts of delay on these two outcomes by lowering the total costs of delay. We hypothesize, however, that the relationship with behavioral intentions will be mediated by attitudes.

Research Methodology

The study was conducted in an experimental setting to control delay, site depth, and familiarity, as well as to allow measurement of outcome variables. Sites were created that covered two levels of each of three factors, providing a $2 \times 2 \times 2$ design with two between-subjects factors (delay and depth) and one within-subjects factor (familiarity). We employed a completely counterbalanced, fully factorial design, providing all 32 combinations of order, delay, depth, and familiarity.[2]

Operationalization of Variables.

Delay. Delay was manipulated with Javascript code to provide a constant eight-second delay per page for the slow site, with no such delay for the fast site. An eight-second delay is about the middle of the range of what others have considered serious (Kuhmann 1989, Ramsay et al. 1998, Shneiderman 1998, Zona Research Study 1999, Hoxmeier and DiCesare 2000), and has been shown to be well past the patience levels of laboratory subjects (Galletta et al. 2004).

A manipulation check for the delay variable asked subjects to evaluate the speed of displaying the Web pages. On a seven-point Likert scale, subjects using the "fast" site, on average, rated the speed as 6.0 (standard deviation 1.2) and subjects on the "slow" site rated the speed as 2.2 (standard deviation 1.3). The difference was significant ($t = 312.3$; one-tailed $p < 0.001$).

[2] To prevent any learning effects in the second task, we counterbalanced completely the order for within-subjects factors. 32 combinations = 2 familiarity orders (familiar–unfamiliar and unfamiliar–familiar) × 4 speed conditions (fast–fast, fast–slow, slow–fast, slow–slow) × 4 depth conditions (deep–deep, deep–broad, broad–deep, broad–broad).

Familiarity. Two artificial websites were created. The "familiar" site contained images, prices, and descriptions of products found in grocery and/or hardware stores almost anywhere, to ensure that subjects would recognize these products and categories of products (for example, a top-level category was "Health Care Products"). In contrast, the "unfamiliar" site contained fictitious software products and accessories, and their categories were meaningless (e.g., a top-level category called "Novo Products" represented the name of a firm or division). Subjects could not make use of previous experience outside the laboratory in searching through the unfamiliar site.

A manipulation check for the familiarity variable asked subjects to evaluate their level of familiarity with the subject matter in the website. Subjects on average rated the familiar site 5.4 (standard deviation 1.4) and unfamiliar site as 2.2 (standard deviation 1.4) on a seven-point Likert scale. The difference was significant ($t = 21.3$; one-tailed $p < 0.001$).

Breadth. The experimental websites were created with two different hierarchical structures for the 81 bottom-level pages. A "broad" type was created with two levels to traverse (nine hyperlinks per page), and a "deep" structure contained four levels to traverse (three hyperlinks per page). These levels were chosen to be consistent with previous studies of interactive system design (Miller 1981, Kiger 1984, Landauer and Nachbar 1985, Norman and Chin 1988).

The websites utilized a linear design where lower child pages were accessible only through their parent page. We did not provide the ability to search or skip levels with shortcuts or other nonlinear methods of navigation. The only exception was a link to the site's home page located on every page to provide a lifeline to a well-known starting point.

A manipulation check for the breadth variable asked subjects to evaluate how they felt about the number of further choices or links on each page visited in the website ($1 = \text{few}$; $7 = \text{many}$). The manipulation check for breadth failed. However, an analysis of the subjects who browsed the unfamiliar site as the second site revealed a significant manipulation check, suggesting that subjects could perceive and judge the depth/breadth manipulation only on the painful unfamiliar site and only after encountering the familiar site as a benchmark.[3]

Table 3 **Average Actual vs. Predicted Clicks for Navigation to a Given Node**

	Deep (4 levels to traverse with 3 links each)	Broad (2 levels to traverse with 9 links each)
Familiar site	6.5 (predicted: 4)	3.8 (predicted: 2)
Unfamiliar site	26.5 (predicted: 40)	8.2 (predicted: 10)

Given that not only perceptions, but also performance is measured in this study, a second manipulation check for breadth was more objective, examining how closely subjects conformed to the expected number of clicks in each treatment cell. We examined all available subject logs[4] to see how closely they conformed to the idealistic numbers suggested by the formula and Tables 1 and 2. Table 3 presents the average actual number of clicks for each condition, which followed the general pattern predicted. The deviations showed that performance was a little worse than expected in the familiar site, and a little better than expected in the unfamiliar site. It is possible that subjects in the familiar site might have missed some of the cues designed to be familiar. It is also possible that participants in the unfamiliar site might have built some on-the-spot familiarity with the completely made-up and meaningless link names as they went along. Subjects in the familiar site might have performed worse than expected because they made random errors, missed the significance of some of the familiar categories, or operated using their own taxonomies that might not have matched ours as completely as we expected.

[3] On a seven-point Likert scale, subjects on average rated the breadth of the broad site 4.1 (standard deviation 1.9) and the deep site 3.9 (standard deviation 1.6). Although these means are in the expected direction, the difference was not significant ($t = 1.0$; one-tailed $p < 0.16$), suggesting that either the condition or the manipulation check measure was inadequate. The next step, performing separate breadth manipulation checks for the unfamiliar and familiar sites was also unsuccessful. However, if the familiar site was encountered first, the breadth of the unfamiliar site was correctly judged to be lower in the deep site (3.9 versus 4.6 for the broad site; $t = 1.7$; one-tailed $p < 0.046$). Judging the familiar site after encountering the unfamiliar site resulted in a failed manipulation check (breadth of 3.55 for the broad site versus 3.45 for the deep site).

[4] Only the first 128 (64 per site) were available due to a malfunction in later data collection sessions.

Table 4 Correlations Among the Measured Variables in the Model

Dependent measure	Scales	Attitudes	Performance	Behavioral intentions
Attitudes	Seven nine-point scales	(0.95)	0.495***	0.774***
Performance	Nine dichotomous items (right or wrong)	0.495***	(0.90)	0.392***
Behavioral intentions	Two seven-point scales	0.774***	0.392***	(0.94)

Notes. Reliabilities on the diagonal.
*** $p < 0.001$.

Dependent Variables. Table 4 provides a summary of the scales on each instrument, instrument reliabilities, and correlations among the three measured items. Multicollinearity was ruled out by examining VIF scores after performing three regressions, making each measure the dependent variable in turn. The largest VIF score was 2.3, well below the common threshold of 4.

Attitudes. Attitudes about the sites were measured by averaging the responses to a set of seven nine-point Likert-type questions adapted from Part 3 of the long form of the QUIS (questionnaire for user interaction satisfaction; Shneiderman 1998, p. 136), which has demonstrated reliability and validity (Chin et al. 1988). Cronbach's alpha was 0.95.

Performance. Performance was operationalized by counting how many of nine[5] search tasks were accomplished. The tasks required subjects to visit each third of the site's hierarchy the same number of times, providing a reasonably balanced overall view of the site. The questions for each site only differed in the search object names (general store products versus software products). Subjects needed to visit exactly the same position of the structure in each site, and the only differences between the familiar and unfamiliar sites were the products and product categories themselves.

The tasks required browsing the site for facts (such as prices or packaging) about various products, and one point was assigned for each correct answer. Short, dichotomous instruments are not normally subjected to reliability analysis, so the results of such analysis should not be viewed with the same expectations as Likert-type instruments or dichotomous instruments with over 20 items (Nunnally 1978). Nevertheless, the statistic of the Kuder-Richardson-20, or KR-20 test (analogous to alpha, but for dichotomous items), was quite high in this study (0.90).

Behavioral Intentions. Behavioral intentions were measured using the average of two questions from Galletta et al. (2004), addressing two related future behaviors: how readily the subject would visit the site again and how likely he or she would recommend that others visit the site (seven-point scales). The alpha score for this very short instrument was also extremely high, at 0.94, as in the previous study.

Subjects

Following a pretest,[6] undergraduate business majors sampled from sections of an upper-division undergraduate MIS course at a large northeastern U.S. university were invited to participate in the study. Of 191 students invited, 177 (93%) volunteered to participate. The first 160 served to fill 5 completely counterbalanced groups of 32 each (77 females, 83 males), and the last 17 were discarded. Subjects were given the opportunity to participate during one of their class periods, and instructors provided extra credit points to stimulate participation. The within- and between-subjects design provided an equal number of subjects (40) for each of the 8 cells.

Subjects were randomly assigned to one of the 32 counterbalanced conditions. In general, ANOVA revealed no differences among the 8 cells on gender, computer efficacy (measured by the instrument from Compeau and Higgins 1995), and computer experience (measured by two scales adapted from Nelson 1991, following the procedure of Hartzel 1999).[7]

[5] We performed hypothesis tests (as shown later in the analysis section) for all tasks separately and found nearly identical results. We also examined the progression of performance from one task to the next, and found that there was no systematic degradation of performance over time.

[6] A pretest employing a separate group of 32 subjects from the same sampling pool (1 for each of the 32 combinations of conditions) in a previous term helped us refine the instruments and procedures. Those data are not included in our results.

[7] Of the 21 $2 \times 2 \times 2$ tests (3 main effects and 4 interactions for each of the three demographic variables), only one difference was significant: computer efficacy differed along a two-way interaction of delay and depth ($F = 4.7$; $p = 0.032$; 1,149 df). One test out of

Procedure

The websites were created and their 32 combinations written on CDs[8] to precisely control the browser's response time. A computer laboratory containing 46 identical Pentium II computers with 17" XGA screens further controlled the subjects' environments. Upon their arrival, participants were asked to sign an "informed consent" form, which stated that participation was voluntary and that they could leave the study at any time. They were then instructed to enter the randomly assigned code number (1–32) printed on their packet to trigger the correct treatment condition. Background logging software provided assurance that the correct number was entered.

A small sample site provided a "warm-up" task, after which participants were prompted to complete the required nine tasks for each of the two main sites. Participants were asked to complete the tasks in order and to write down the answers to the questions as they progressed. Answers were assessed as correct or incorrect by comparing subjects' answers against the facts on the site. Data about the actual number of clicks were recorded automatically. At the conclusion of each main site, subjects were instructed to close the browser window and complete the questions addressing attitudes and behavioral intentions. On average, subjects completed the experiment in about an hour.

Our site was typical of most websites that group product pages into intermediate categories linked together from a top page, and the tasks typify common questions for which users might seek answers. Every effort was made to keep the sites and tasks as realistic as possible.

Results

As per Neter et al. (1996), we began our analysis with a multivariate ANOVA for both attitudes and performance, followed by a univariate ANOVA for each (using SPSS version 10). Because each MANOVA test was significant, univariate tests were run. Tests of mediation of attitudes between the experimental factors and behavioral intentions were also conducted, the first according to Baron and Kenney (1986) and the second with PLS.

Table 5 Cell Means for Attitudes

	Slow		Fast	
Treatment	Mean	Standard deviation	Mean	Standard deviation
Unfamiliar				
Deep	1.8	1.0	2.8	1.5
Broad	2.5	1.4	4.0	2.1
Familiar				
Deep	4.8	1.9	6.4	1.6
Broad	6.2	1.7	6.6	1.2

Tables 5 and 6 provide means for each dependent variable, and Table 7 provides the multivariate tests of Hypothesis 1 and Hypothesis 2. The tests reveal that each three-way interaction was significant ($p < 0.002$) as predicted. Therefore, univariate testing was performed for the hypotheses.

Individual ANOVA tests were run for both outcome variables, and the amount of variance explained by each analysis was 0.557 for attitudes and 0.498 for performance (adjusted R^2).

Means for the dependent variables across the main treatments are presented in Tables 8–10. Tests of main effects could serve as another category of manipulation checks, or at least to supplement the large body of monofactor research. All 6 tests are significant (in the direction exhibited in previous studies) at the 0.001 level, demonstrating that each experimental factor had the intended and expected effects on both dependent variables. It is notable that the Depth/Breadth

Table 6 Cell Means for Performance

	Slow		Fast	
Treatment	Mean	Standard deviation	Mean	Standard deviation
Unfamiliar				
Deep	0.42	0.29	0.80	0.30
Broad	0.81	0.23	0.93	0.17
Familiar				
Deep	0.95	0.09	0.98	0.05
Broad	1.0	0	1.0	0.02

21 is expected to be significant at the 0.05 level based purely by chance, and a Bonferroni adjustment renders it nonsignificant.

[8] The CD, as well as all instruments and materials, are available from the first author on request.

Table 7 Overall MANOVA: Attitudes and Performance

Factor	*F*	Sig
Delay	37.3	0.000
Familiarity	216.0	0.000
Breadth	35.8	0.000
Delay ∗ Familiarity	17.2	0.000
Delay ∗ Breadth	6.4	0.002
Depth ∗ Familiarity	15.6	0.000
Delay ∗ Familiarity ∗ Breadth	7.7	0.001

treatment shows strong results as opposed to the results we obtained on our manipulation checks. That is, while participants had difficulty assessing their Depth/Breadth treatment upon reflection, the actual performance and attitude results of that treatment were both strong.

Hypothesis Testing: Direct Effects on Attitudes and Performance

Statistically speaking, supporting each hypothesis would require a three-way interaction among the experimental factors on both attitudes and performance. Such an interaction would demonstrate that the dampening effect of familiarity and breadth on delay are enhanced more for a slower site than for a faster site. The results support both hypotheses, demonstrating a three-way interaction for both attitudes ($F = 5.9$; $p < 0.016$) and task performance ($F = 7.98$; $p < 0.005$). The cell means reveal that familiarity and breadth interact to reduce the performance and attitudinal disadvantage of a slow site. The special provision of Hypothesis 2, that attitudes serve as a mediator toward behavioral intentions, requires additional analysis, which will be examined next.

Hypothesis Testing: Mediating Role of Attitudes Toward Behavioral Intentions

Hypothesis 2 states that the experimental effects would be manifested through attitudes as a mediator. This model could not be tested directly because the independent variable in the model was actually three dichotomous factors. Therefore, two approaches were used. A mediation test using regressions was followed by structural equation modeling using PLS, after adapting the data to fit the technique. Neither test is completely adequate by itself because the mediation test does not take into account the entire model, and the PLS model cannot include the experimental factors separately.

Following the procedure of Baron and Kenney (1986), a test of mediation was performed. To establish mediation, (a) the independent variable must correlate with the mediator variable, (b) the independent variable must affect the dependent variable in a regression of the independent variable on the dependent variable, (c) the mediator must affect the dependent variable when regressing both the independent and the mediator variables on the dependent variable, and (d) the effect of the independent variable on the dependent variable in (c) must be less than in (b).

In this case, we did not have a single independent variable but a set (the three experimental factors). Therefore, in all three steps, we used the three dichotomous independent variables in the regressions. Table 11 provides the statistics for each test. Indeed, the effects of the independent variables (delay, familiarity, and breadth), on behavioral intentions, the dependent variable, were less in (c) than in (b), and attitudes therefore appear to mediate the experimental effects as hypothesized.

Table 8 Means Across the Delay Treatment

	Slow		Fast		
Dependent measure	Mean	Standard deviation	Mean	Standard deviation	Difference (F, p) at 1,312 df
Attitudes	3.8	2.3	5.0	2.3	$F = 42.1$; $p < 0.001$
Performance	0.80	0.30	0.92	0.19	$F = 40.8$; $p < 0.001$

Table 9 Means Across the Breadth Treatment

	Deep		Broad		
Dependent measure	Mean	Standard deviation	Mean	Standard deviation	Difference (F, p) at 1,312 df
Attitudes	3.9	2.4	4.8	2.3	$F = 27.3$; $p < 0.001$
Performance	0.79	0.31	0.93	0.16	$F = 52.0$; $p < 0.001$

Table 10 Means Across the Familiarity Treatment

	Unfamiliar		Familiar		
Dependent measure	Mean	Standard deviation	Mean	Standard deviation	Difference (F, p) at 1,312 df
Attitudes	2.8	1.7	6.0	1.8	$F = 331.7$; $p < 0.001$
Performance	0.74	0.31	0.98	0.06	$F = 144.0$; $p < 0.001$

Table 11 Mediation Test Statistics

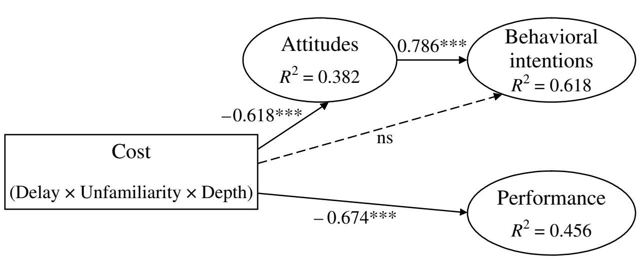
	Step (a): Regression of cost on attitudes	Step (b): Regression of factors alone on behavioral intentions	Step (c): Regression of factors plus attitudes on behavioral intentions
Adjusted R^2 (overall)	0.564	0.396	0.571
Coefficients for experimental factors for Steps (b) and (c):			
—Delay t value		3.2***	2.5***
—Familiarity t value		13.0***	2.4***
—Breadth t value		3.2***	0.3***

***$p < 0.01$.

For the PLS approach, we reduced the three dichotomous factors to a single one that we called "cost" by multiplying them as the three-way interaction would imply. A code of one was assigned for the "low-cost" conditions: fast, familiar, and broad, and a code of two was assigned for each of the costly conditions: slow, unfamiliar, and deep. The resultant independent variable was similar to the inverse of scent, where the worst of all conditions would receive an 8 ($2 \times 2 \times 2$), the best would be scored as 1 ($1 \times 1 \times 1$), and other cells would be somewhere in between.

PLSGraph 2.91.03.04 was employed to examine both the model hypothesized (with attitudes as a mediator), and a model with attitudes, behavioral intentions, and performance as separate outcomes. Figure 2 provides the results of the model that was tested as specified, and Figure 3 examines an alternative model without attitudes as a mediator. The hypothesized model provides a much higher R^2 for behavioral intentions, and the explained variance for attitudes was nearly unchanged. When paths from both attitudes to behavioral intentions and cost to behavioral intentions are included, the cost to behavioral intention path is not significant, suggesting full mediation.

Figure 2 PLS Model with Hypothesized Paths



Note. The "ns" path was not hypothesized but is included for completeness.
***$p < 0.001$.

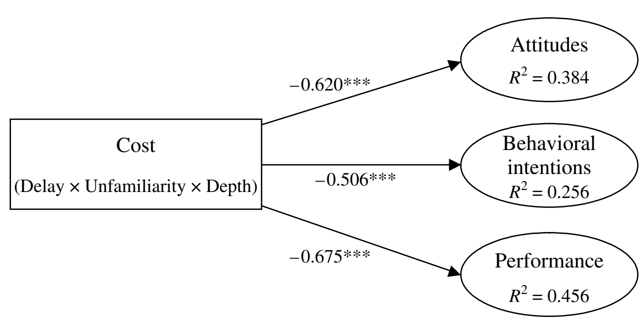
Our results provide evidence that behavioral intentions are indeed mediated by attitudes under our conditions, and other models may exist for which attitudes should not be omitted. Thus, the TAM approach of removing attitudes from models leading to behavioral intentions, which has become common practice over the years, should be applied only with caution.

In summary, we performed an experiment with manipulations that appeared to work as desired, given our successful manipulation checks and powerful main effects of website delay, breadth, and familiarity on user attitudes and performance. Such effects are consistent with previous monofactor literature. Tests of the model provided support for our hypotheses that providing a high degree of information scent through short delay, familiarity, and breadth combine to predict both user performance and attitudes. Attitudes, in turn, act as a mediator in predicting behavioral intentions.

Discussion

Our model predicted a dampening effect of familiarity and breadth on the negative effects of delay on users' attitudes and performance on nine search tasks; and

Figure 3 PLS Model Without Attitudes as a Mediator



***$p < 0.001$.

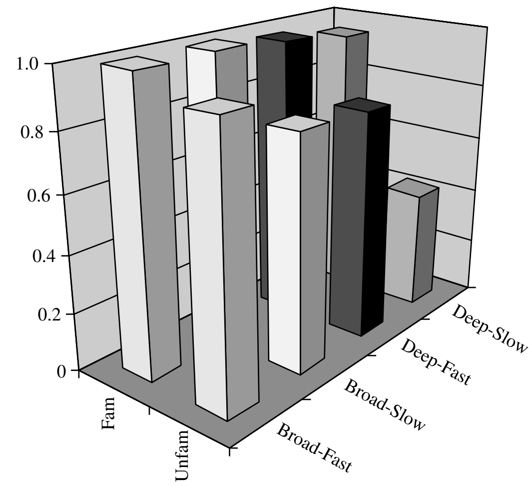
predicted that attitudes would mediate the impact of the three factors on behavioral intentions. The interaction goes beyond the previous literature, which largely studied only main effects. Our study attempted to provide a richer view of affective and performance outcomes of delay and the other cost factors.

The three-way ANOVAs explained 56% of the variance in attitudes and 50% of the variance in performance. Although the ANOVA model's main effects were consistent with the previous literature, the model goes beyond main effects, predicting that ill effects of delay are dampened by providing familiar structure and content and a broad site with few levels. Although it is relatively rare to theorize and find three-way interactions, the effects can be seen graphically in Figures 4 and 5. The shortest bar in each bar chart represents the lowest performance (Figure 4) and the least positive attitudes (Figure 5), respectively. Figure 4 implies that our familiarity and breadth treatments were similar in impact on performance, while Figure 5 seems to credit more of the dampening effect to the familiarity treatment on attitudes. More formal assessments are in order.

Comparing the *F* ratios[9] in Tables 9 and 10 will demonstrate that the effect of familiarity on performance actually appears to be 2.8 times stronger than the effect of breadth on performance, while the effect of familiarity on attitudes is 12.1 times stronger than the effect of breadth on attitudes. It is possible that our familiarity treatment was perhaps stronger than intended given that there was guaranteed to be no familiarity with the terminology at all, or our breadth treatment was weaker than intended given that the depth decreased only from five levels to three levels. Our choice of the familiarity was driven by the desire to minimize variance due to individual differences, and to be sure there would be no familiarity in the unfamiliar site among any subjects. Our choice of the number of levels was based on prior literature in menu design.

Although future studies could examine how to better "equalize" the treatments, it is important at this stage at least to recognize that there is a clear dampening effect from both manipulations, evidenced by

Figure 4 Three-Way Interaction for Performance

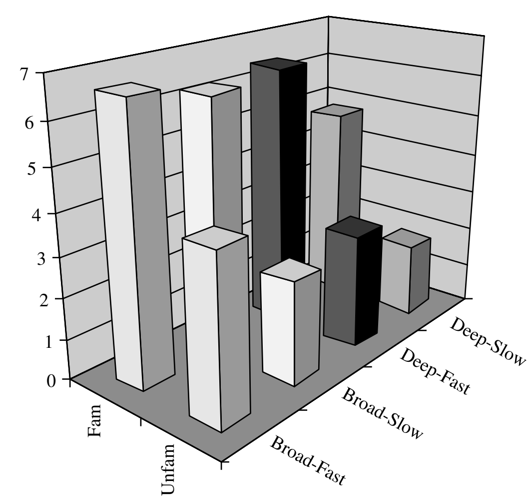


a strong three-way interaction. Examination of the ANOVA results and the bar charts seem to indicate that all three factors provide the richest understanding of delay and account for the maximum variation in performance and attitudes.

Limitations

As in most studies, there are several limitations that should be kept in mind before generalizing from the results of this study. First, the subjects were college students, and the results might not generalize to the rest of the population. Fortunately, Voich (1995) found students to be particularly representative of values and beliefs of individuals employed in a variety of

Figure 5 Three-Way Interaction for Attitudes



[9] For the comparisons, the *F* ratios were divided as follows: performance 144.0/52.0; attitudes 331.7/27.3.

occupations. Also, we did not expect students to react differently to the stimuli than would other individuals. Thus, the materials were designed to tap rather "invariant" (Simon 1990) activities such as Web use.

Another limitation is that the task might not represent typical tasks that online shoppers undertake. To address this potential problem, we attempted to make the tasks as realistic as possible. It is perhaps common to find shoppers browsing through a site, looking for facts and details of products that they might buy. Our task required precisely this activity. The main departure from traditional usage was the lack of "search" facilities. A search option would have destroyed our control over the paths followed and would have confounded our results. In addition, for a variety of reasons described earlier, a search facility is not a silver bullet for finding information on a website.

Our objective to separate the effects of delay, familiarity, and breadth on the outcome variables forced us to employ a highly structured, symmetrical website that conformed to a rigid hierarchy. In the real world, it is rare to find such ideally modeled sites when doing casual information searches. In our case, this choice was made to remove any confounding effects of variations in structure. Fortunately, many product-oriented sites are indeed hierarchical and structured, especially given the number of sites that are driven by database retrievals to provide the content. Although perfect symmetry is rare, all or nearly all menu design studies have made use of symmetrical hierarchies so that conclusions can be drawn about a particular choice of breadth. Future research should compare nonsymmetrical designs to symmetrical ones.

Analysis was limited by our inability to use structural equation modeling to test the complete model (Gefen et al. 2000), in one simultaneous analysis. The discrete nature of the treatments required use of two statistical procedures for mediation testing. However, combining the three experimental factors into a composite factor enabled us to examine a structural model. Although the more detailed ANOVA results addressed the experimental treatments more precisely, the structural model provided a method for assessing the relationships throughout the entire model.

Finally, incentives of laboratory browsers and home shoppers might differ. In short, home shoppers can choose a different store if they become frustrated searching for a desired product, while our users did not desire any of the products but participated to receive extra credit. Although extra credit is indeed an incentive, it does not reflect incentives of actual shopping. Tracking real shoppers on real online stores would have provided greater realism, but the cost would be a loss of control. Our sacrifice provided greater assurance that the effects we found were caused by the treatments. We might have actually muted our findings because students in a lab with a set of tasks placed in front of them would not be as likely to care as much about the task or be as likely to suffer reduced performance (quitting before their tasks were all complete).

Implications and Conclusions

This study is meant to provide several implications for both researchers and practitioners. Interaction findings augment two decades of investigation that were limited mostly to main effects. The three factors are synergistic because they operate in multiplicative, rather than additive, mode given our ANOVA results. As such, one or more of the factors can be manipulated to compensate for difficulties in the other factors. That is, if the scent is low because of a distant target many levels away, performance and attitudes can be improved by reducing the delay and/or increasing the familiarity of the links. Likewise, if the scent is low due to low familiarity and long delay, the site can be broadened.

Therefore, rather than focusing only on the impact of any one of the factors on attitudes or performance, it is important to consider the level of the other two factors. Galletta et al. (2004) discovered the importance of familiarity in examining a regression line for performance, attitudes, and behavioral intentions. Regressions of outcomes based on different levels of delay were different for familiar and unfamiliar sites—there was less patience for unfamiliar sites. This study adds depth and breadth to the mix.

Studying outcomes of performance, attitudes, and behavioral intentions together provide a more complete story than studying only one of the outcomes. Although additional antecedents make models more precise, additional outcomes make the model more useful by explaining variance in additional outcome variables with the same experimental manipulations.

Future studies might examine outcomes such as time, stress, recall, confidence, flow, or actual purchasing behavior.

The scope of the analysis is another issue of importance to researchers. Successful investigation of the three factors in a business-to-consumer environment provides some promise for future studies not only in this environment, but also in organizational applications such as an intranet site or a corporate portal. Extending our results to other factors or other settings will help to better understand the model's scope and applicability. Will the relationships hold in an intranet, in an entertainment site, or in a site that holds a great deal of intrinsic motivation?

Information foraging theory appears to be a promising basis for research in Web design. Users appear to have strong reactions as the costs of browsing for information increase, possibly exceeding the benefits. Furthermore, time is a potent component of foraging cost, but only considering delay time would omit other important factors such as familiarity and depth, two factors that add to foraging cost. These factors allow a researcher to perform the more complete analysis that is suggested by estimating foraging cost with a formula that predicts the nature of the users' likely path to a goal and the distance to that goal. That is, Foraging Theory can be enhanced by considering delay along with familiarity and breadth.

Although our model appears to explain a large amount of variation in attitudes and task performance, there is ample opportunity to focus on other antecedents that might lead to greater understanding of the usability of websites. Although we were able to explain variance in performance and behavioral intentions with our model, other studies should also include additional factors such as motivation, perceived need, or screen size (using various types of devices such as PDAs or cell phones) that may help to better understand these relationships under different circumstances.

Practitioners should also benefit from this study. They should be advised to pay close attention to each of the three individual factors within their control to the greatest extent possible. Page-loading time should be minimized, links to bottom-level pages should use the most familiar possible terminology, and intermediate pages should have more—rather than fewer—links per page. Although minimizing page-loading time is often an explicit goal of designers, less is stated about the depth of a site, and even less about the terminology that guides users to the target node of a site. Although familiarity may be more subjective and difficult to measure, this study shows that it is quite potent in the determination of scent and resultant cost of browsing.

More importantly, however, although practitioners are not usually made aware of interactions among experimental factors, the interactions in this study provide some clear, pragmatic advice for site design: First and foremost, if delay and depth appear to be problems, designers should pay even more attention to the terminology used to categorize Web content. This research suggests that designers should be careful when deviating from established industry norms or commonly used terms. In those cases when designers do choose more site-specific terminology, user testing to help determine the familiarity of site content and terminology becomes even more important. Only testing can uncover future problems (Gould and Lewis 1985). One promising tool to help test a design is the cognitive walkthrough (Blackmon et al. 2003), where users speak aloud while attempting to perform a task. Weaknesses will become obvious when users become lost, and any terminology causing those difficulties can be noted and improved.

When it is difficult or impossible to expect a reasonable amount of familiarity in target users, then a great deal of effort should be expended to compensate. Designers can maximize breadth by providing many links per page to limit users' trial-and-error browsing. Also, delay can be minimized by eliminating unnecessary graphic content; streamlining page design; or assuring proper server configuration, redundancy, and/or bandwidth. When the site is broadened and made faster in loading, users suffer less cost in browsing.

Designers can also be less concerned about delay if familiarity is expected to be high and the site has few levels. In those cases, pages can be made more attractive by making use of additional graphics or more extensive underlying applications. Alternatively, the site can be made with greater depth than might otherwise have been considered advantageous. Speed can be considered as a compensatory mechanism for factors such as depth and lack of familiarity.

In some situations, heavy graphical content might be required on a site, or it may be anticipated that it will be difficult to keep up with the demand for the site. In those cases, breadth should be maximized and the site's structure should employ very familiar, clear terminology.

Although it is useful in the evolution of our understanding to consider design factors one at a time, such as those investigated in previous studies, analysis of interactions provide the basis for more thorough understanding for researchers and more practical heuristics, and choices, for designers.

Delay, familiarity, and breadth of interactive systems have been studied separately for the last few decades. This study establishes that the three design factors are not independent and should be studied together. In an experimental study of 160 undergraduate subjects, the three factors were effective in predicting user performance in, and attitudes about, navigating a shopping site. Further, attitudes fully mediated the effect of the factors on behavioral intentions. Future studies should build on such interactive models rather than only on main effects from the factors taken separately.

Acknowledgments

The authors would like to thank Enrique Mu, Dong-Gil Ko, and Alex Lopes for allowing their students to participate. The authors also appreciate the excellent comments received on previous versions of this paper from the attendees of the Pre-ICIS HCI Workshop and presentations at the University of Illinois at Urbana-Champaign, University of Kansas, Duquesne University, and Temple University. The authors also thank the reviewers, the Associate Editor, and the Senior Editor for their detailed and helpful suggestions.

References

Ajzen, I. 2001. Nature and operation of attitudes. *Annual Rev. Psych.* **52** 27–58.

Barber, R. E., H. C. Lucas. 1983. System response time, operator productivity, and job satisfaction. *Comm. ACM* **26**(11) 972–986.

Baron, R. M., D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personality Soc. Psych.* **51**(6) 1173–1182.

Bernard, M. L. 2002. Examining the effects of hypertext shape on user performance. *Usability News* **4**(2). http://psychology.wichita.edu/surl/usabilitynews/42/hypertext.htm.

Blackmon, M. H., M. Kitajima, P. G. Polson. 2003. Repairing usability problems identified by the cognitive walkthrough for the Web. *Proc. SIGCHI*, Ft. Lauderdale, FL. Association for Computing Machinery, 497–504.

Butler, T. W. 1983. Computer response time and user performance. *Proc. SIGCHI*, Boston, MA. Association for Computing Machinery, 58–63.

Card, S. K., T. P. Moran, A. Newell. 1983. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Card, S. K., P. Pirolli, M. Van Der Wege, J. B. Morrison, R. W. Reeder, P. K. Schraedley, J. Boshart. 2001. Information scent as a driver of Web behavior graphs: Results of a protocol analysis method for Web usability. *Proc. SIGCHI*, Seattle, WA. Association for Computing Machinery.

Chae, M., J. Kim. 2003. An empirical study on the breadth and depth tradeoffs in very small screens: Focusing on mobile internet phones. *Amer. Conf. Comput. Inform. Systems*, Tampa, FL. Association for Information Systems, 2083–2093.

Chau, P. Y. K., G. Au, K. Y. Tam. 2000. Impact of information presentation modes on online shopping: An empirical evaluation of a broadband interactive shopping service. *J. Organ. Comput. Electronic Commerce* **10**(1) 1–22.

Chi, E. H., P. Pirolli, J. Pitkow. 2000. The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. *Proc. SIGCHI*, The Hague, The Netherlands. Association for Computing Machinery, 161–168.

Chi, E. H., P. Pirolli, K. Chen, J. Pitkow. 2001. Using information scent to model user information needs and actions on the Web. *Proc. SIGCHI*, Seattle, WA. Association for Computing Machinery.

Chin, J. P., V. A. Diehl, K. L. Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. *Proc. SIGCHI*, Washington, D.C. Association for Computing Machinery, 213–218.

Chiravuri, A., L. Peracchio. 2003. Investigating online consumer behavior using thin slices of usability of web sites. *Amer. Conf. Comput. Infom. Systems*, Tampa, FL. Association for Information Systems, 2105–2110.

Commerce Department. 2005. Quarterly retail e-commerce sales: 1st quarter 2005. U.S. Department of Commerce News (May 20, 2005). http://www.census.gov/mrts/www/ecomm.html.

Compeau, D. R., C. A. Higgins. 1995. Computer self-efficacy: Development of a measure and initial test. *MIS Quart.* **19**(2) 189–211.

Dannenbring, G. L. 1983. The effect of computer response time on user performance and satisfaction: A preliminary investigation. *Behav. Res. Methods Instrumentation* **15**(2) 213–216.

Davis, F. D. 1989. Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quart.* **13**(3) 319–340.

Dellaert, B. G. C., B. E. Kahn. 2000. How tolerable is delay? Consumers' evaluation of Internet websites after waiting. *J. Interactive Marketing* **13**(1) 41–59.

Dumais, S. T., T. K. Landauer. 1983. Using examples to describe categories. *Proc. SIGCHI*, Boston, MA. Association for Computing Machinery, 112–115.

Edwards, D. M., L. Hardman. 1989. "Lost in hyperspace": Cognitive mapping and navigation in a hypertext environment, Ch. 7. R. McAleese, ed. *Hypertext: Theory into Practice*. Ablex Publishing, NJ, 90–105.

Fishbein, M., I. Ajzen. 1975. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA.

Frese, M. 1987. A theory of control and complexity: Implications for software design and integration of the computer system into the work place. M. Frese, E. Ulich, W. Dzida, eds. *Psychological Issues of Human-Computer Interaction at the Work Place*. North-Holland, Amsterdam, The Netherlands, 313–337.

Galletta, D., R. Henry, S. McCoy, P. Polak. 2004. Web site delays: How slow can you go? *J. Assoc. Inform. Systems* **5**(1) 1–28.

Gauch, S., J. B. Smith. 1991. Search improvement via automatic query reformulation. *ACM Trans. Inform. Systems* **9**(3) 249–280.

Gefen, D., D. W. Straub, M. C. Boudreau. 2000. Structural equation modeling and regression: Guidelines for research practice. *Comm. Assoc. Inform. Systems* **4**(7) 1–77.

Gould, J. D., C. Lewis. 1985. Designing for usability: Key principles and what designers think. *Comm. ACM* **28**(3) 300–311.

Hartzel, K. S. 1999. Understanding the role of client characteristics, expectations, and desires in packaged software satisfaction. Unpublished doctoral thesis, Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA.

Hoxmeier, J. A., C. DiCesare. 2000. System response time and user satisfaction: An experimental study of browser based applications. *Amer. Conf. Comput. Inform. Systems*, Association for Information Systems, Long Beach, CA, 140–145.

Jacko, J. A., G. Salvendy. 1996. Hierarchical menu design: Breadth, depth, and task complexity. *Perceptual Motor Skills* **82**(3) 1187–1202.

Jacko, J. A., G. Salvendy, R. J. Koubek. 1995. Modeling of menu design in computerized work. *Interacting Comput.* **7**(3) 304–330.

Johnson, E. J., J. W. Payne. 1985. Effort and accuracy in choice. *Management Sci.* **31**(4) 395–414.

Katz, M. A., M. D. Byrne. 2003. The effect of scent and breadth on use of site-specific search on e-commerce web sites. *ACM Trans. Human-Comput. Interaction* **10**(3) 198–220.

Kiger, J. 1984. The depth/breadth trade-off in the design of menu-driven user interfaces. *Internat. J. Man-Machine Stud.* **20** 201–213.

Koblentz, E. 2000. Web hosting partner keeps stone site rolling. *Eweek* (October 30) 27.

Korthauer, R. D., R. J. Koubek. 1994. An empirical evaluation of knowledge, cognitive style, and structure upon the performance of a hypertext task. *Internat. J. Human-Comput. Interaction* **6**(4) 373–390.

Kuhmann, W. 1989. Experimental investigation of stress-inducing properties of system response times. *Ergonomics* **32**(3) 271–280.

Landauer, T., D. Nachbar. 1985. Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width. *Proc. SIGCHI*, New York. Association for Computing Machinery, 73–78.

Larson, K., M. Czerwinski. 1998. Web page design: Implications of memory, structure and scent for information retrieval. *Proc. SIGCHI*, Los Angeles, CA. Association for Computing Machinery, 25–31.

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

Lee, A. T. 1999. Web usability: A review of the research. *SIGCHI Bull.* **31**(1) 38–40.

Lee, E., J. MacGregor. 1985. Minimizing user search time in menu retrieval systems. *Human Factors* **27** 157–162.

Lohse, G. L., P. Spiller. 1998. Electronic shopping. *Comm. ACM* **41**(7) 81–87.

Martin, G. L., K. G. Corl. 1986. System response time effects on user productivity. *Behav. Inform. Tech.* **5**(1) 3–13.

McKinney, V., K. Yoon, F. Zahedi. 2002. The measurement of Web-customer satisfaction: An expectation and disconfirmation approach. *Inform. Systems Res.* **13**(3) 296–315.

Miller, D. 1981. The depth/breadth tradeoff in hierarchical computer menus. *Proc. 25th Annual Meeting Human Factors Soc.*, Human Factors Society, New York, 296–300.

Miller, R. B. 1968. Response time in man-computer conversational transactions. *Proc. Fall Joint Comput. Conf.*, AFIPS, San Francisco, CA, 267–277.

Muramatsu, J., W. Pratt. 2001. Transparent queries: Investigation users' mental models of search engines. *Proc. 24th Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval*, New Orleans, LA. Association for Computing Machinery, 217–224.

Nah, F. 2002. A study of Web users' waiting time. V. Sugumaran, ed. *Intelligent Support Systems Technology: Knowledge Management*. IRM Press, Hershey, PA, 145–152.

Nah, F. H., K. Kim. 2000. World wide wait. M. Khosrowpour, ed. *Managing Web-Enabled Technologies in Organizations: A Global Perspective*. Idea Publishing Group, Hershey, PA, 146–161.

Nelson, R. R. 1991. Educational needs as perceived by IS and end-user personnel: A survey of knowledge and skill requirements. *MIS Quart.* **15**(4) 503–525.

Neter, J., M. H. Kutner, C. J. Nachtsheim, W. Wasserman. 1996. *Applied Linear Statistical Models*, 4th ed. Irwin Publishers, Chicago, IL.

Nielsen, J. 1997. The need for speed. http://www.useit.com/alertbox/9703a.html.

Nielsen, J. 1999a. Who commits the "top ten mistakes" of Web design? http://www.useit.com/alertbox/990516.html.

Nielsen, J. 1999b. The top ten *new* mistakes of Web design. http://www.useit.com/alertbox/990530.html.

Norman, K., J. Chin. 1988. The effect of tree structure on search in a hierarchical menu selection system. *Behav. Inform. Tech.* **7**(1) 51–65.

Nunnally, J. 1978. *Psychometric Theory*. McGraw-Hill, New York.

Olston, C., E. H. Chi. 2003. ScentTrails: Integrating browsing and searching on the Web. *ACM Trans. Comput.-Human Interaction* **10**(3) 177–197.

Otter, M., H. Johnson. 2000. Lost in hyperspace: Metrics and mental models. *Interacting Comput.* **13** 1–40.

Paap, K. R., R. J. Roske-Hofstrand. 1988. Design of menus. M. Helander, ed. *Handbook of Human-Computer Interaction*. Elsevier, Amsterdam, 205–235.

Pardue, J. H., J. P. Landry. 2001. Evaluation of alternative interface designs for e-tail shopping: An empirical study of three generalized hierarchical navigation schemes. *Amer. Conf. Comput. Inform. Systems*, Association for Information Systems, Boston, MA, 1335–1337.

Parkinson, S., M. Hill, N. Sisson, C. Viera. 1988. Effects of breadth, depth, and number of responses on computer menu search performance. *Internat. J. Man-Machine Stud.* **28** 683–692.

Pirolli, P., S. Card. 1999. Information foraging. *Psych. Rev.* **106**(4) 643–675.

Pitkow, C., J. Kehoe, J. Rogers. 1998. GVU's 9th WWW user survey. Georgia Institute of Technology, Atlanta, GA. http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-04/reports/.

Polak, P. 2002. The direct and interactive effects of web site delay length, delay variability, and feedback on user attitudes, performance, and intentions. Unpublished doctoral dissertation, Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA.

Ramsay, J., A. Barbesi, J. Preece. 1998. A psychological investigation of long retrieval times on the World Wide Web. *Interacting Comput.* **10**(1) 77–86.

Rose, G. M., D. Straub. 2001. The effect of download time on consumer attitude toward the e-service retailer. *e-Service J.* **1**(1) 55–76.

Rose, G., M. Khoo, D. W. Straub. 1999. Current technological impediments to business-to-consumer electronic commerce. *Comm. AIS* **1**(Article 16) 1–74.

Rose, G. M., J. Lees, M. Meuter. 2001. A refined view of download time impacts on e-consumer attitudes and patronage intentions toward e-retailers. *Internat. J. Media Management* **3**(2) 105–111.

Sawasdichair, N., S. Poggenpohl. 2002. User purposes and information-seeking behaviors in Web-based media: A user-centered approach to information design on websites. *Sympos. Designing Interactive Systems. Proc. Conf. Designing Interactive Systems: Processes, Practices, Methods, and Techniques,* London, England. ACM Press, New York, 201–212.

Schwartz, J. P., K. L. Norman. 1986. The importance of item distinctiveness on performance using a menu selection system. *Behav. Inform. Tech.* **5**(2) 173–182.

Sears, A., J. A. Jacko. 2000. Understanding the relation between network quality of service and the usability of distributed multimedia documents. *Human-Comput. Interaction* **15**(1) 43–68.

Sears, A., J. A. Jacko, E. M. Dubach. 2000. International aspects of WWW usability and the role of high-end graphical enhancements. *Internat. J. Human-Comput. Interaction* **12**(2) 243–263.

Selvidge, P. 2003. Examining tolerance for online delays. *Usability News* **5**(1). http://psychology.wichita.edu/surl/usabilitynews/51/delaytime.htm.

Shneiderman, B. 1998. *Designing the User Interface*. Addison-Wesley, Boston, MA.

Simon, H. A. 1990. Invariants of human behavior. *Annual Rev. Psych.* **41** 1–19.

Somberg, B., M. Picardi. 1983. Locus of the information familiarity effect in the search of computer menus. *Proc. 27th Annual Meeting Human Factors Soc.*, Norfolk, VA, 826–830.

Song, J., F. Zahedi. 2001. Web design in e-commerce: A theory and empirical analysis. *Proc. Internat. Conf. Inform. Systems*, Session T3.3. New Orleans, LA. Association for Information Systems, 205–220.

Taylor, S., P. A. Todd. 1995. Understanding information technology usage: A test of competing models. *Inform. Systems Res.* **6** 144–176.

Thuring, M., J. Hannemann, J. M. Haake. 1995. Hypermedia and cognition: Designing for comprehension. *Comm. ACM* **38**(8) 57–66.

Todd, P., I. Benbasat. 1999. Evaluating the impact of DSS, cognitive effort, and incentives on strategy selection. *Inform. Systems Res.* **10**(4) 356–374.

Torkzadeh, G., G. Dhillon. 2002. Measuring factors that influence the success of internet commerce. *Inform. Systems Res.* **13**(2) 187–204.

Turban, E., D. Gehrke. 2000. Determinants of e-commerce website (sic). *Human Systems Management* **19** 111–120.

Venkatesh, V., M. G. Morris, G. B. Davis, F. D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quart.* **27**(3) 425–478.

Voich, D. 1995. *Comparative Empirical Analysis of Cultural Values and Perceptions of Political Economy Issues*. Praeger, Westport, Connecticut.

Wonnacott, L. 2000. When user experience is at risk, Tier 1 providers can help. *Infoworld* (November 6) 76.

Zaphiris, P., L. Mtei. 1997. Depth vs. breadth in the arrangement of Web links. University of Maryland, College Park, MD. http://www. otal.umd.edu/SHORE/.

Zona Research Study. 1999. The need for speed. Press release. http://www.useit.com/alertbox/990530.html.