Effects of Scent and Breadth on Use of Site-Specific Search on E-Commerce Web Sites

MICHAEL A. KATZ and MICHAEL D. BYRNE
Rice University

Users faced with Web sites containing many possibly relevant pages often have a decision to make about navigation: use the menu of links or use the provided site search function. Two studies were conducted to examine what users do when faced with this decision on e-commerce Web sites, and how users go about deciding which method to attempt. An exploratory study revealed a wide distribution of searching and browsing behavior across sites and users. Counter to some predictions, use of the site search functions did not yield faster or more accurate performance in locating products. Questionnaire data suggested that factors relevant to the menu structure, interface element prominence, information scent and user dispositions all influenced the decision of whether to browse or search a site for a product. A second experiment utilizing novel e-commerce sites and allowing for more control of factors found to be important in the first study found that browsing behavior was influenced by both the breadth and information scent of the menus. These results suggest that providing site search should not be used to compensate for poor menu design, and provide further evidence regarding the design of effective menu structures.

Categories and Subject Descriptors: G.4 [**Mathematical Software**]—*User Interfaces*; H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*Navigation*

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Searching, browsing, WWW, navigation, information retrieval, empirical studies, e-commerce

1. INTRODUCTION

When a Web site contains only a few pages, navigation is generally a fairly trivial process for the user. However, as sites grow in page count, the demands placed on the user in terms of navigation grow. In response to this, many sites now provide a "local" search function, which will search the contents of only that site. Unfortunately, empirical evidence regarding the impact this has on the

Authors' addresses: Michael A. Katz, Yahoo! Inc., 701 First Avenue, Sunnyvale, CA 94809; email: mikekatz@yahoo-inc.com; Michael D. Byrne, Department of Psychology, Rice University, 6100 Main Street MS-25, Houston, TX 77005; email: byrne@acm.org.

navigation task is, at best, sparse. When faced with the choice between browsing the site by traversing the hyperlinked menus or using the local search, what do users actually do? Equally importantly, how do they decide which course of action to take? These are the primary issues of concern in this paper.

1.1 What Do Users Do?

Although companies exist that track the number of online users and Web sites (such as MediaMetrix), there is relatively little data available regarding the use of local search facilities. More specifically, many sites offer users the choice between browsing a (usually hierarchical) menu structure and a site-specific search facility. To be clear about the domain of inquiry, in this context we take "search" to mean the use of a site-specific search function accessed by typing a search query into a text field. We are not referring to use of search engines or indices such as Google or Yahoo! In contrast, we take "browse" to mean alternate methods of navigating a particular Web site, particularly the traversal of the site's menu hierarchy. Note that we do not take "browse" to mean activity that is not goal-directed, but take it as an alternative to searching.

Empirical studies aimed at understanding how and when users choose to search vs. browse are few and lack depth. This raises a number of empirical questions for which there are few published attempts at answers:

- Is searching faster than browsing? Is it more accurate?
- What proportion of users search and what proportion browse?
- Is there an interaction of user and site that influences the decision to search or browse?
- How often do users begin with one strategy but then switch to the other?

Research with hypertext systems that predate the present-day World Wide Web has attempted to address similar questions albeit for particular information retrieval systems (e.g., Rada and Murphy [1992]; Campagnoni and Ehrlich [1989]; Marchionini [1989]). The results of one such study [Rada and Murphy 1992] suggested that different hypertext systems were better suited to either search or browse tasks.

Turning to the Web, many researchers are willing to make claims regarding search behavior and prevalence. Scanlon [2000] claims that users rely on the search engine approximately half of the time, while Nielsen [2000] argued that approximately half of users are "search-dominant." Perhaps driven by a belief in the obvious face validity of such claims, many sites provide search functions, and e-commerce sites do so with regularity. Powell [2000] suggests that sites with regularly formatted data, complex-to-digest data, or sites that contain more than 100 pages should include a local search engine.

Despite the many claims regarding the motivation for a search function, the reasons users choose such a function, and the existence of patterns of searching behavior, there is surprisingly little empirical data used to support these claims. Many researchers make such statements based solely on the subjective experience of observing users on the Web. For example, Powell [2000] claims that searching is a faster means of finding information than browsing, but evidence

exists to suggest that this is often not true [Scanlon 2000]. Furthermore, can users truly be characterized as search-dominant? Can situations (or sites) be found where the search-dominant user will consistently browse and vice-versa [Sawyer 2000]?

Of course, there are clear limitations in the empirical detail of the literature just discussed. While there have certainly been rigorous, peer-reviewed empirical studies of Web behavior, particularly dealing with the tasks undertaken [i.e., Byrne et al. 1999; Sellen et al. 2002] and link-following based on some measure of "information scent," [i.e., Card et al. 2001; Blackmon et al. 2002], none have focused on the questions surrounding the use of site search facilities. The current research is intended to provide empirical evidence relevant to many of these issues, with the ultimate goal of informing Web site design with respect to these questions. One of the larger segments of the present Web environment are e-commerce sites, so these are the focus of study. This is also a domain in which many site designers may want users to browse rather than search for a variety of reasons. While providing local site search may be technically straightforward and intuitively appealing, it may not in fact be the best choice.

1.2 How Do Users Decide What to Do?

One approach to investigating user behavior on e-commerce Web sites is to determine the factors involved in a user's decision to use a menu or site search function to locate products. Two theoretical perspectives are examined that can potentially account for the decision processes of the user who must choose between searching and browsing on e-commerce sites.

1.2.1 *The Cost-Benefit Perspective.* One approach to explaining human decision-making is through a cognitive cost-benefit analysis, involving the consideration of various positive and negative dimensions of alternative strategies for a task [Kleinmuntz and Schkade 1993; Payne et al. 1993]. This approach focuses primarily on two cost-benefit dimensions: the cognitive effort required to use a strategy and the ability of a strategy to produce an accurate response. Individuals select the particular strategy that represents the best accuracy-effort tradeoff, or more appropriately, anticipated accuracy-effort tradeoff [Kleinmuntz and Schkade 1993] for the task at hand. In the context of how individuals gather information, information foraging theory argues that people modify their strategies to maximize their rate of gaining valuable information while minimizing cost, with "cost" referring to the costs of accessing, rendering, and interpreting information-bearing items [Pirolli and Card 1995, 1999]. The decision faced by Web surfers who seek items on a site that provides local search functionality is the problem of where to allocate search time: to the search function or to the menu hierarchy?

A derivative of foraging theory, "information scent" [Pirolli and Card 1995; Pirolli 1997] describes the amount of remote information a user can derive regarding the location of information based on the design or labeling of the information structure. Computational modeling of human information foraging has suggested that users' browsing choices are based on the evaluation of information scent [Pirolli and Card 1999], and this approach has been codified for

use as an evaluation tool [Chi et al. 2000]. However, this account has not been extended to situations in which the user has the option of a site-specific search.

Of course, query formation and document retrieval have been studied prior to the advent of the Web. The notion of information scent bears some similarity to the notions of "topicality" or "relevance" as discussed in the document retrieval literature (e.g., Schamber et al. [1990]), though that literature has also been concerned with a host of other factors such as perceived authority and quality (see Wang and Soergel [1998], for a review). Because the retrieval of library-type documents is a problem with much larger scope than seeking specific information on a single Web site, the extent to which such additional factors play a role in this context is not clear. For this problem, we believe information scent is the appropriate theoretical perspective.

In the context of shopping online, information scent could refer to the amount of information a user could attain regarding the location of a product in a site based solely on the design of the site's home page. For instance, category labels (serving as product headings) on a site can be more or less distinctive resulting in differing degrees of information scent [Larson and Czerwinski 1998; Tilson et al. 1998], with distinctiveness referring to the semantic aspects of an alternative that enhances its difference from other alternatives in the set [Norman 1991]. Users attempt to find the best semantic match between the task description (or goal) and the available labels on display objects (such as menus and tool bars [Soto 1999].

Presumably, information scent can play a large role in a user's decision to traverse the menus or use the site search function. A site with poor information scent (due to uninformative product categories) should result in the user anticipating browsing to be a more costly strategy than searching. The information foraging and scent models generally index scent with a measure of semantic similarity, which seems like a reasonable starting point (see Pirolli and Card [1999] for a discussion. Note that Blackmon et al. [2002] also use semantic similarity in a similar fashion). This account should predict that, all other things being equal, rate of site search use should be a function of the semantic similarity of the category labels to the item being sought.

Equally compelling as information scent to the user's cost-benefit analysis of a site is the breadth and depth of the product headings (i.e., product categories). Note that "breadth" refers to the number of options at a level in a hierarchy, while "depth" refers to the number of levels in a hierarchy. In fact, the two issues are related in that hierarchies with general top-level categories tend to require a greater number of lower levels. A review of breadth versus depth research suggested that breadth was favored over depth in most studies of the organization of menu contents [Larson and Czerwinski 1998]. With increased breadth, there are increased demands on visual search processes along with corollary increases in time required to make a decision. Conversely, with increased depth, there are more decisions required by the user, and perhaps more importantly, there is greater uncertainty as to the location of target items. Although a deep menu may require less processing time than a broad menu for each level of the hierarchy, this advantage is often outweighed by the greater number of hierarchy levels in a deep menu [Norman 1991].

As an example relevant to online retailers, a site that presents only two product categories on its home page, namely "Men's Clothes" and "Women's Clothes" will imply to the user that many more levels of the product hierarchy will necessarily be covered before a sought after product is located. Conversely, a site with many top level product categories such as "Scarves," "Hats," "Blouses," and "Pants" implies to the user that relatively few levels of the hierarchy remain to be explored (see also Lohse and Spiller [1998]).

Despite more recent research, the debate of breadth versus depth in information structures has yet to be resolved. While prior studies tend to support the virtues of high breadth, important points have been made regarding the disadvantages of such an approach. The current study will seek to further clarify this debate and provide data in the realm of e-commerce sites.

1.2.2 *The Attentional Capture Perspective.* Properties of the visual environment may be major determinants of whether an object in the environment draws (or "captures") attention. For instance, unique objects in terms of color or brightness have been found to capture attention in a visual display, as have objects that lie on or near visual boundaries [Todd and Kramer 1993]. The attentional capture perspective predicts that the prominence of various display elements can be a critical factor in determining users' choice of browsing or searching. It has been shown in other domains that various information display factors can bias decision-making [Payne et al. 1993], and this should apply to Web sites as well.

Initial orientation to a new environment is supported by landmarks [Vinson 1999; Evans 1980; Golledge et al. 1985]. If a search function is perceived as a landmark by the user due to brightness contrast of the text field with the background of the page, then the user may perceive alternative strategies for finding a product to be subordinate (such as traversing the product hierarchy). In fact, a strategy based on visual dominance may be adaptive; page designers may be sending an implicit message about the value of the search function by making it difficult to visually locate.

The purpose of Study 1 was twofold. An attempt was made to characterize product-finding behavior on e-commerce Web sites and to determine factors involved in the decision to use a search function or browse a menu to find a product. To achieve this end, participants were instructed to locate items on e-commerce sites and questionnaires were used to provide a window into the participants' decision making processes. The factors discussed above provided by the cost-benefit and attentional capture perspectives offer potential explanations of user behavior. A main goal of this study was to determine the extent to which these factors are in fact relevant to the product-finding behavior of users.

2. STUDY 1

2.1 Method

2.1.1 *Participants.* Twenty Rice University undergraduate students participated to fulfill a course requirement. As part of the "decision" questionnaire

Fig. 1. Deerskin.com.

administered for each site in set 2 (see Materials below), participants were asked whether they had used the experimental sites previously. They reported having previously used a mean of 1.5% of the sites.

Following the experiment, participants were asked to fill out the "user profile" questionnaire (see Materials below). Of the twenty participants, all but two reported more than five years of general computer usage, and all but one reported having used the Web at least two hours per week for more than one year. Regarding current Web usage, all but two participants reported at least five hours per week. Sixteen participants reported purchasing items online in the past, with eleven reporting having purchased at least five items online.

2.1.2 *Materials.* Twenty e-commerce sites were selected with the intention of eliciting a wide range of searching and browsing behavior among participants. Such an assessment was made informally by the first author based on the cost-benefit and attentional capture principles discussed earlier. Two such sites are presented in Figures 1 and 2, namely deerskin.com and bluefly.com. Any site that did not have both browsing and searching capability was excluded from the pool of eligible sites. (Note that many of the sites may have been redesigned since the study.)

For each site, we selected a broad sample of items available on the site for use in the study. The number of items selected equaled the number of participants in the study (20) so that each participant was asked to find a different item for each site. This requirement was meant to minimize the effect of the items chosen on user behavior; we were interested in user behavior for particular sites, not particular items. Furthermore, as the sites used covered a broad range of industries, it was not possible to ask the participants to locate the same items

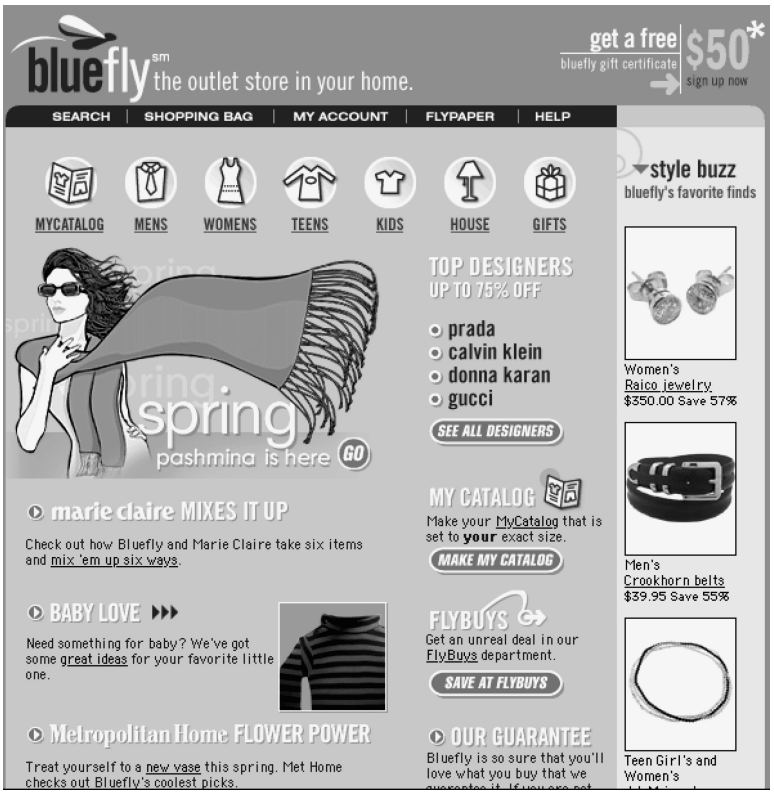


Fig. 2. Bluefly.com.

for each site. In fact, such a strategy might have produced a confound as participants could be influenced by previous success rates at locating a particular item using a particular method.

Two questionnaires were also developed to determine the factors involved in the decision to search or browse. The first questionnaire, the "decision" questionnaire, consisted of 20 questions intended to reveal the identity and role of important factors involved in the decision process; the questions were based on the cost-benefit and attentional capture factors discussed earlier. Questions were developed with the aim of eliciting both interpretations of available information as well as predictions from participants as to what they expected to see once they proceeded. For example, two of the questions were as follows: "After clicking on the menu option you would choose, how many other menus would you expect to have to go through to find the product?" and "On a scale of 1 (low) to 10 (high), rate how clear you are as to what you should type in the Search field to find the item." Furthermore, two versions of the decision questionnaire were created to accommodate participants who chose to search and those who chose to browse. The second questionnaire, the "user profile" questionnaire, was designed to obtain general user data regarding Web experience and typical online behavior patterns.

A Sony DCR-TRV10 Digital Camcorder was used to videotape the experimental sessions. Each participant was tested individually on an Apple iMac personal computer running Microsoft Internet Explorer 4.5 with an Ethernet/T1 Internet connection. The digital video (with audio) from each participant was captured onto an Apple G4 computer and analyzed with the task analysis software MacSHAPA [Sanderson et al. 1994].

2.1.3 *Design.* Twenty sites were used. Each participant received two sets of sites, each set containing ten unique sites, with set 1 always preceding set 2. For half of the participants, ten of the twenty sites were assigned to set 1 with the other ten assigned to set 2. For the remaining half of participants, the assignment of sites to the two sets was reversed. In other words, the sites used in set 1 for half of the participants were the sites used in set 2 for the other half. This counterbalancing scheme was necessary as a different procedure was used for set 1 and set 2 (see Procedure below).

Ten of the twenty items for each site met the requirement that only one item existed for that item specification (e.g., "Air Storm GPT Softball bat") while the other 10 item specifications were satisfied by a maximum of 20 items available on the site (e.g., "Softball bat"). The two sets of ten were matched pairs as in the example so that the "specific" item would satisfy the "general" item requirement (i.e., "Air Storm GPT Softball bat" is one of the available "Softball bats").

Each participant received twenty unique items, one unique item for each of the twenty sites, with the following requirements. Half of the items presented to each participant were general and half specific (as defined above) and, over all participants, half of the items sought for each site were general and half specific.

2.1.4 *Procedure.* The sites were presented in a random order to each participant, and the item to be found on each site was randomly assigned to each participant (in accordance with the above requirements). The items to be found by the participants were presented one at a time on separate sheets of paper (one item on one sheet for each individual site), and participants were instructed that the purpose of the experiment was to investigate how people find products on e-commerce sites.

For the first set of ten sites, the participant located the item presented to her and wrote down the price of the product on the sheet of paper used to present the item. The experimenter recorded whether the user used the menus or site search function for the site as this was the dependent variable of primary interest. If both behaviors were exhibited, the experimenter noted the sequence of participant behaviors. Finally, if the participant wished to cease looking for a particular item, the experiment was continued with the next item.

For the second set of ten sites, the participant was shown the item to be located on the sheet of paper as in the first set. The participant then wrote down on the sheet of paper used to present the item what she decided to do to locate the product. As an example, she might have written, "I would click on the 'Men's Clothes' menu option." Then the decision questionnaire was administered to the participant. During administration of the questionnaire, the site was left

visible and the participant was permitted to scroll (but not click on) the site. Upon completion of the questionnaire, the participant was instructed to actually locate the item and provide the price information (as in the first set). Following this cycle of events, the next site was loaded by the experimenter and the cycle restarted. Following both sets of stimuli, the user profile questionnaire was administered to the participant.

In the interest of gaining further insight into potential factors influencing searching and browsing behavior, the participants were instructed to "think aloud" while looking for the items and the entire experimental session was videotaped.

As mentioned earlier, the main dependent variable of interest was whether the participants' initial attempt at locating each product was via browsing or searching. Success rate data (for locating products) and time to complete the task were also recorded. Participants were tested between March 23 and April 10, 2000, so it is important to note that many of the experimental sites may have been re-designed since the testing period.

2.2 Results

Participants were very accurate in locating items. After all attempts to locate products, the overall accuracy rate was 94%. An analysis presented later describes accuracy rates for individual attempts. Five (1.3%) items were missing at the time of testing, five (1.3%) items were located on the homepage of the site, two (<1%) sites would not load, and one (<1%) site crashed in progress. For within-subjects statistical tests, missing data were replaced with the grand (search/browse) mean. Once the possibility of missing items was realized (due to site inventory changes), the presence of all future items was verified prior to further testing. As a result, 23 (5.8%) items were replaced with new items. For the two cases in which a site would not load, a replacement site was used (stressless.com) but the resulting data from that site was not included in any analyses due to the infrequent occurrence of such failures.

Although more specific items were predicted to lead to a greater likelihood of searching behavior, such a pattern was not sufficiently consistent across sites to reach significance, $t(19) = -1.07$, $p = 0.30$. The overall search rate for general items (across sites) was 0.42 while the overall search rate for specific items (across sites) was 0.45.

2.2.1 *Frequency and Time Analyses.* The frequency of searching and browsing behavior and corresponding time spent on each activity was recorded and analyzed, and success rates were also noted. The current analyses include all browsing or searching attempts for a particular item and therefore the overall accuracy rate reported earlier (94%) does not apply to this section (that success rate measured the accuracy of product locations after all attempts).

For this analysis, a browse attempt was defined as an attempt to locate a product without the use of the search function, and the attempt ended when participants either found the item (success), quit looking for the item, switched strategies (and began using the search function), backtracked to a higher menu level, or otherwise made a clear change in course of action (such as selecting

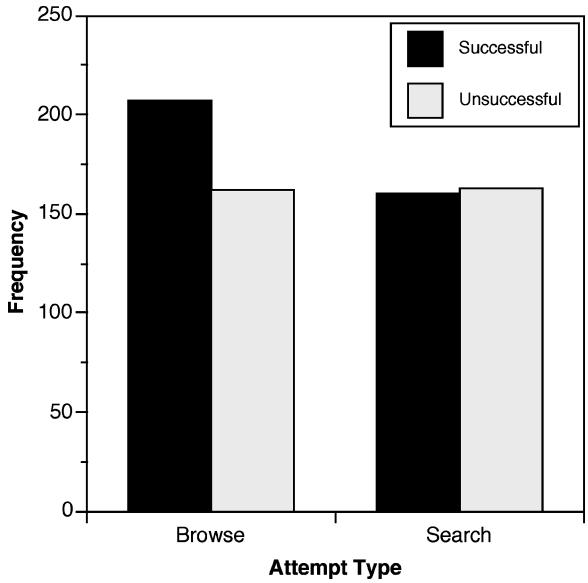


Fig. 3. Attempt frequency as a function of success and type.

an unrelated item on the site home page). A search attempt was defined as an attempt to locate the item using the search function and such an attempt ended when participants either found the item (success), quit looking for the item, switched strategies (and began browsing the menus), made a new search attempt, or otherwise made a clear change in course of action.

The global frequency and time data (collapsing across all sites; Figures 3 and 4 respectively), clearly indicate that the use of the search function in an effort to locate a product did not lead to a greater rate of success. Although it failed to reach statistical significance, participants showed a higher success rate when using the menus to find items as opposed to search, $\chi^2(1, \mathrm{N} = 692) = 2.98$, $p = 0.08$ (see Figure 5).

Similarly, and perhaps more counterintuitively, searching was also not reliably faster than browsing. A more clear predictor of time was in fact whether the attempt was successful or not, as depicted in Figure 4.

A 2×2 ANOVA with success (i.e., successful/unsuccessful) and product location method (i.e., browse/search) as independent factors and time as the dependent measure yielded no effect of product location method, $F < 1$. Further, no significant interaction was found, $F < 1$. In contrast, the success effect was highly reliable, $F(1, 73) = 11.10$, $p = 0.001$, with successful product locations taking more time than unsuccessful attempts. This finding is less surprising when one considers that participants would often immediately come to the conclusion that their chosen course of action would not be effective, and then frequently either search again or backtrack to a different menu option.

Participants often made several attempts to find an item with the mean number of attempts per item being 1.75. However, it was found that participants were willing to switch strategies (i.e., browsing to searching and vice-versa)

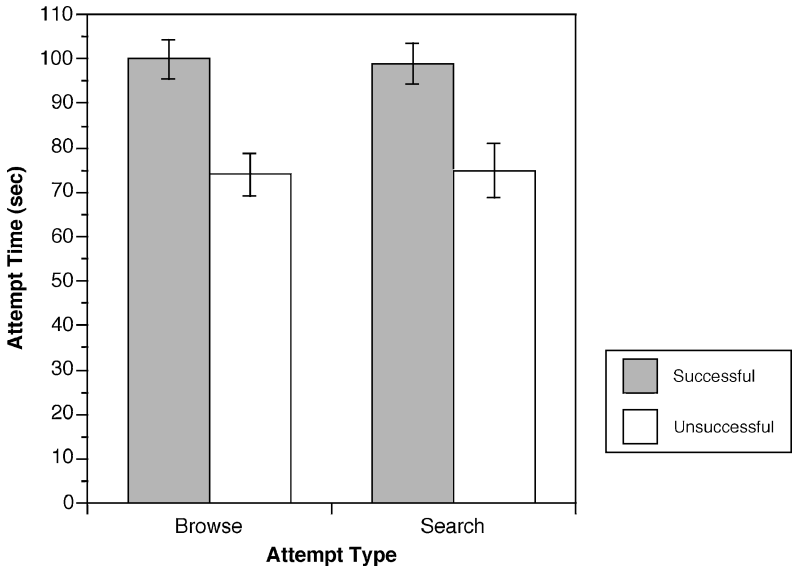


Fig. 4. Attempt duration (with standard error) as a function of type and success.

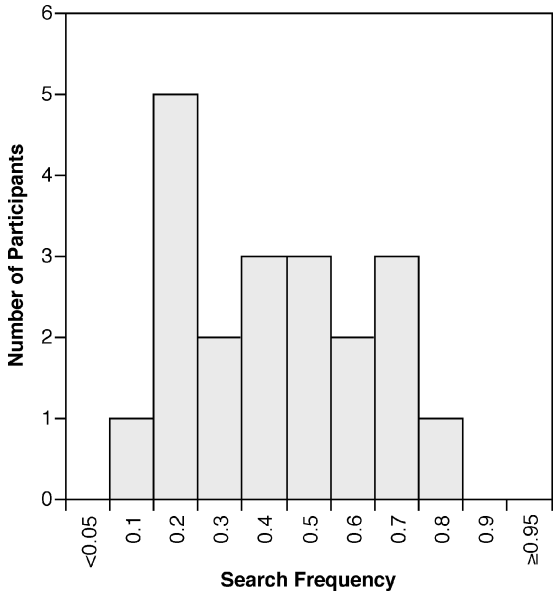


Fig. 5. Search frequency histogram by participant (bins are .10 wide).

when they felt the current strategy would not prove fruitful. It should be noted, however, that switching was not the only course of action when an approach was deemed ineffective. On many occasions, for example, participants would backtrack to a higher menu level and select another path through the menu hierarchy. In addition, it was discovered that participants would often switch strategies several times on a single site indicating that the decision on how to

Table I. Reliable Predictors of Initial Search According to Logistic Regression Procedure

Logistic Regression Predictor	Coefficient	*p*-Value
How often use search function to find products when shopping	4.99	<0.001
Considered time as a factor in searching vs. browsing	2.19	<0.001
How long will it take to find products using menus	1.31	<0.001
Number of menus expect to have to go through to find product	0.32	0.008
How clear as to what should type into search field	0.32	0.020
How noticeable is search option on page	0.22	0.040
How many items would you expect search to return	−0.04	0.020
How noticeable is menu of options on page	−0.48	<0.001
How long will it take for search to find item	−0.74	0.005

locate an item is a fluid and dynamic process. The mean number of switches while looking for each item was 0.31, but if one focuses only on those cases for which participants switched at least once, the mean number of switches per item rises to 1.58.

2.2.2 *Questionnaire Analyses.* Factors like the predicted efficacy and prominence of the menu and search function were investigated with the "decision" questionnaire. These data, along with response data from the user profile questionnaires were used as predictors in a logistic regression to determine which factors influenced the strategic decision on how best to locate an item.

Product location behavior (i.e. search/browse) was regressed on the more than 50 quantitative variables collected from the questionnaires. Using stepwise predictor selection, a model was produced that yielded several significant predictors (see Table I).

Examples of factors that failed to reach significance were "whether there existed an appropriate number of menu options (on the home page)" and "the number of items expected to be found on the page containing the item."

Free response answers from the questionnaires were then sorted into meaningful categories and the frequencies of such responses were correlated with the search rates from set 2 by site. This way, data could be obtained regarding the factors that participants believed were involved in their decision to search or browse. An example of a qualitative question from the decision questionnaire was "Did you consider time as a factor in choosing to use the Search function versus browse the menus? *If yes, in what way*? " [italics added] and a sample answer (to the follow-up question) was "Search engine is easier."

To conduct such an analysis, responses for each qualitative question were placed into the most appropriate category (or categories) and the proportion of responses in each category was tallied for each site. So, for the above example, the response was sorted into a category labeled "Search is easier/faster/more direct." Then, the proportion of responses to the question "Did you consider time as a factor in choosing to use the Search function versus browse the menus? *If yes, in what way*? [italics added]" that could be placed into the category "Search is easier/faster/more direct" was calculated for each site. For Valueamerica.com, the proportion of responses to that question that fit into that category was 30%. Finally, the response rates for each category (for each question) were correlated with search rates across sites.

Table II. Correlations Between Search Rate and Frequency of Comments

Factor Predicting Search Behavior	r	p-Value
Search is easier/faster/more direct	0.55	0.010
Menu options are vague/incomplete	0.54	0.010
Menu is highly noticeable due to good contrast with background	−0.50	0.020
Search may not work or will be confusing	−0.51	0.020
Good match of item and menu	−0.57	0.009

When the response rates for each response category were correlated with search rates (across sites), several meaningful correlations consistent with the quantitative questionnaire findings were found (see Table II). As a whole, these correlations suggest that the opinions and expectations of participants played a key role in their decision of whether to search or use the menus to find a product. More importantly, the perceived effect of their opinions and expectations were consistent with the actual effect of the relevant factors.

2.2.3 *Search Rates Across Participants and Sites.* Another set of empirical questions concern the extent to which sites and users can be said to be search-dominant or menu-dominant. To answer this, we looked at the initial product location attempt for each site and user. There was a highly significant main effect of site on search rates, $F(19, 361) = 3.55, p < 0.001$, yet an attempt to cluster the sites based on prevalence of search usage yielded no clear-cut groupings. It is especially interesting to note that only four sites had a search rate greater than 50%.

Search rates also varied by participants with some participants content to search on most sites but others reluctant to do so unless absolutely necessary. Figure 5 represents the search rate distribution by participant. There does not appear to be a clear division between "searchers" and "non-searchers" in this sample.

2.3 Discussion of Study 1

Results from our sample of participants and sites revealed that searching behavior varied substantially across users and sites with such behavior failing to lead to faster or more accurate performance when compared to site menu usage. Users were clearly sensitive to a variety of factors when choosing whether to search or browse and their interaction with the sites was heavily influenced by the predicted effort required and the expected likelihood of success of each method.

Together, the quantitative and qualitative analyses provide converging evidence regarding the rationale behind users' decisions to search or browse to find products. Similar results were found pertaining to the factors that govern the decision (quantitative analysis) and the factors that users perceive to be important to the decision (qualitative analysis of user comments). It is fairly well-established (e.g. Nisbett and Wilson, [1977]) that the factors driving decisions and those reported to drive decisions need not agree; however, they seem to in this case.

First, users do appear to be driven by explicit cost-benefit considerations. Users considered time an important factor in their decision-making process,

and thus factors like site structure (e.g., expected number of traversals needed, search function efficacy) also play a role. In addition, factors related to the graphic design of the sites accompanied the influence of the (perceived) underlying information structure. For instance, display prominence led to page elements—such as the search function—capturing user attention and likely played a role in whether that option was used. However, the use of a prominent interface element also likely depends on the information scent of that option relative to other options.

Additional factors were pertinent to the perceived or inferred efficacy of the search function itself as users were (not surprisingly) unwilling to use functionality that they deemed unreliable. A final set of factors represented the user's disposition and preferences, with users possessing a priori opinions of the efficacy of various information gathering options.

The suggestive evidence of Study 1 provided support for the cost-benefit and attentional capture perspectives. However, a more rigorous investigation of these factors is warranted before any firm conclusions should be drawn. This is the motivation for Study 2.

3. STUDY 2

Study 1 suggested that the decision of a user to search or browse a site for a particular item is affected by multiple factors including the user's inclination to search as well as the design and structure of the site. Since site designers cannot directly influence the users' individual characteristics, we choose to focus on those factors which can be affected by site design decisions: the structure and layout of the site. Users indicated that they did consider factors such as perceived label relevance and number of menu levels. Study 2 was thus conducted to investigate the effects of menu breadth and information scent on e-commerce Web site browsing and searching behavior in a more controlled environment.

While some evidence bearing on these questions was obtained in Study 1 regarding the influence of breadth and scent on searching behavior, the materials used in that study contained many uncontrolled elements. As actual company Web sites were used, menu breadth and information scent could not be systematically varied and many other elements of the sites such as the fonts, colors, and images used differed among them.

In Study 2, novel e-commerce Web sites were created that varied only on the factors of interest, namely menu breadth and scent. All other elements of the sites were held constant including the presence of images and general layout.

Further, information scent was investigated with a more controlled method known as Latent Semantic Analysis (LSA; described below) and menu breadth was defined in terms of the number of options present. Using these controlled materials, we sought to further investigate our hypotheses put forth in Study 1. Specifically, we sought to determine the influence of information scent and menu breadth on the likelihood to use menus or a search function on e-commerce sites.

3.1 Method

3.1.1 *Participants.* Thirty-two Rice University undergraduate students participated to fulfill a course requirement.

3.1.2 *Materials.* Sixteen e-commerce Web sites were created, which varied the design of the site home pages. The created sites represented every combination of breadth (high/low) and scent (high/low). There were four sites for each condition resulting in a total of sixteen sites.

The created sites were portrayed as sixteen competing sites in the genre of "clothing retailer" (see Figure 6 below for a sample site from each condition). Therefore, the content of the sites was equivalent (to the extent allowed by the varying experimental conditions).

The Web sites contained various elements to allow variation among them while simultaneously maintaining a sense of realism. All sites possessed a product menu, search capability, and eight unique JPEG images of models that were each 100 pixels $\times$ 100 pixels in area. The fonts used were varied among sites (at least among company name headings) and alignment of company names at the top of the page was varied as well. Each site contained a unique company name and color scheme, and five irrelevant (but live) links such as "Customer Service" or "Security Policy." Scrolling was not required for any of the sites.

Accompanying the sixteen Web sites were sixteen matched pairs of items to be located by participants, with each pair consisting of a general and specific version of an item. For example, the general version of an item might have been "sweater" while the specific version might have been "Kenneth Cole Diamond Pattern Sweater." All items were very popular and therefore familiar to all participants (i.e., there was no doubt as to the comprehension of "sweater"). The item specificity manipulation is beyond the scope of this paper but is included here for sake of completeness in describing the experimental design.

Each participant was tested individually on an Apple iMac personal computer running Microsoft Internet Explorer 5 with the display set to a resolution of 1024 $\times$ 768 pixels. The created sites were located on the computer's hard drive so no Internet connection was necessary.

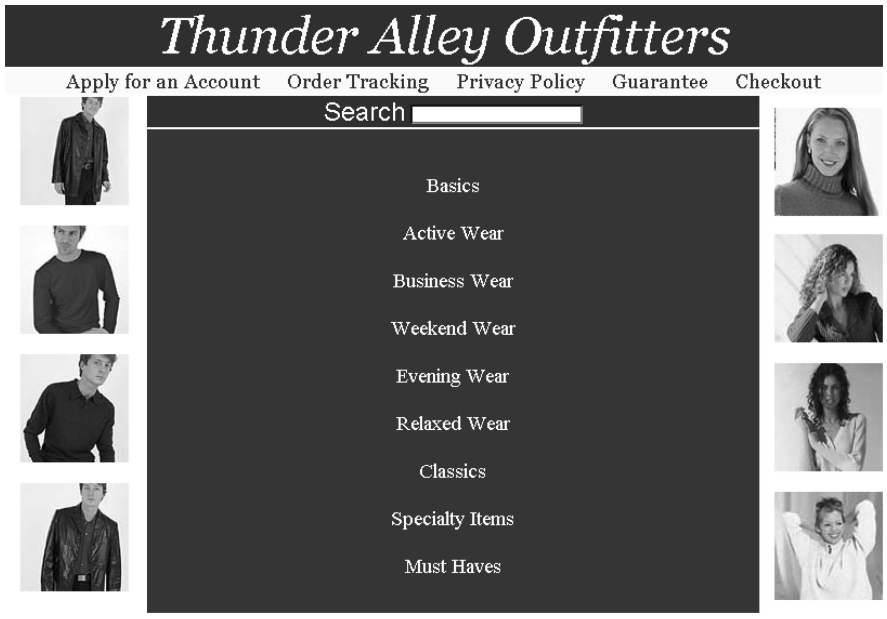
3.1.3 *Design.* The breadth and scent factors were manipulated completely within subjects. Below is a depiction of the factors that were manipulated, the two levels for each factor, and the method of operationalizing each level. Menu breadth was simply either low (9 items) or high (30 items). Information Scent cannot be operationalized quite as easily. While others have measured scent by asking users to rate how confident they are before they click on a link [Spool and Klee 2000], the information scent measure used here was based on a method of measuring semantic similarity known as Latent Semantic Analysis or LSA [Deerwester et al. 1990; Landauer et al. 1998]. LSA is a mathematical technique for extracting and representing the similarity of meaning of words and passages by analysis of large bodies of text. It assumes that the information about all the word contexts in which a word does and does not appear provides a set of mutual constraints that largely determines the similarity of the meaning of words to a set of words [Landauer et al. 1998].

(a)

(b)

Fig. 6. (Panels a–d): Sites used in Study 2. Panel a: Site with high breadth, high scent menu. Panel b: Site with high breadth, low scent menu. Panel c: Site with low breadth, high scent menu. Panel d: Site with low breadth, low scent menu.

(c)



(d)

Fig. 6. Continued.

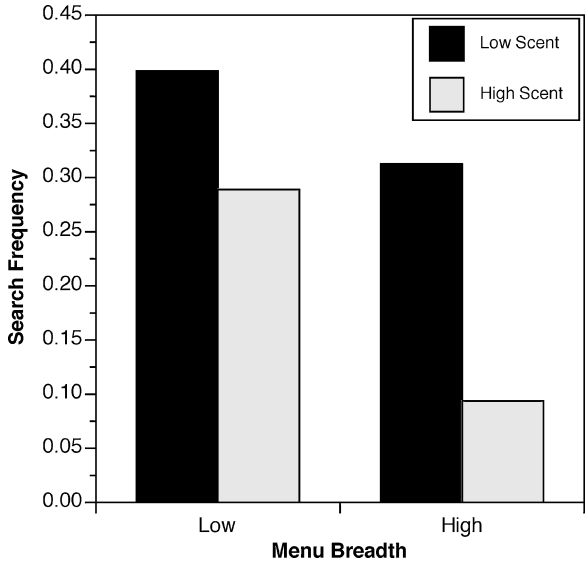


Fig. 7. Search rate as a function of scent and breadth.

The specific method used is based on applications provided by the Science and Applications of Latent Semantic Analysis Group (SALSA) at the University of Colorado at Boulder (see <http://lsa.colorado.edu/>). LSA processes a large sample of language and represents the words as points in a very high dimensional space, the semantic space. Every word can then be represented as a vector, and terms (words or strings of words) represented by summing all the vectors of the constituent words. When two terms are compared, the cosine of the angle between the vectors representing the terms is compared within a given semantic space. Highly-related terms have large cosines.

Using LSA, the following procedure was used to compute the scent measures that were used to quantify the distinction between high scent and low scent menus. First, the cosine between each target item and each menu label was determined for all menu labels and all target items. Second, for each target item, the cosine of the "best" and "worst" menu labels for each of the four main menus was determined.

For example, consider the target item "sweater" and the broad-high scent menu in Figure 7a. This menu contains multiple labels including the most-related term "Sweaters" and the least-related term "Scarves." The mean cosine for all best matches was computed across all menus and target items, as well as the mean cosine for all the worst matches. The mean of the best matches is the "maximum relatedness" measure and the mean of the worst matches the "minimum relatedness" measure.

Broad high-scent menus had a maximum relatedness of 0.55 with a minimum of −0.05. Narrow high-scent menus had a maximum relatedness of 0.46 with a minimum of 0.02. The corresponding numbers for low-scent were 0.38 maximum and −0.07 minimum for broad menus and 0.36 maximum with −0.04 minimum for narrow menus. Note that both high-scent menus and low-scent

menus contained items with low relatedness; the difference is that the high-scent menus had better best-matches.

The target items were counterbalanced so that each target item was sought on every site across participants and each site was viewed only once by each participant. Due to the fact that there were twice as many targets as sites and participants were shown each site only once, four participant groups were required. The participants were assigned to one of two main groups (Group A and Group B) and one of two sub-groups (Sub-group I and Sub-group II). For Group A, sites 1–8 were assigned general items and sites 9–16 were assigned specific items. For Group B, sites 1–8 were assigned specific items and sites 9–16 were assigned general items. The two sub-groups for each main group were necessary to ensure that every item was sought for every site across all participants. For sub-group I, sites 1–8 were assigned items 1–8 while for sub-group II, sites 1–8 were assigned items 9–16.

The dependent measure was again search rate.

3.1.4 *Procedure.* Participants were shown each site one at a time in random order and were presented with a single item provided on a sheet of paper for each site. However, the participants were not actually asked to find the products at any time. After a participant was shown the product to be found, she was instructed to decide what she would do first to locate the item and then perform that action. After performing that action, she was presented with a Javascript message box that simply stated what action she took (e.g., "You clicked on: Pants"). Then, the experimenter recorded the menu option clicked or the text typed into the search field and loaded the next site. The dependent variable of primary interest was the initial course of action taken by participants to use the menu or search function to locate the product.

The reason behind this somewhat unnatural instruction was the lack of functionality behind the sixteen created sites. It would have been a massive undertaking to create sixteen sites with full product menu structures and, more importantly, fully operational search functions. If a participant were to learn that the search functions were non-functional, for example, it could have influenced her decision whether to use the search function on future sites.

3.2 Results

Search rate as a function of scent and breadth is presented in Figure 7. As expected, higher scent yielded less searching, as did broader menus. Both main effects were reliable ($F(1, 31) = 27.59$, $p < 0.001$ for breadth and $F(1, 31) = 12.31$, $p < 0.01$ for scent). The data suggest that the breadth effect is larger for high-scent menus (or perhaps the scent effect is larger for high-breadth menus), but the interaction was merely suggestive, $F(1, 31) = 2.64$, $MSE = 0.145$, $p = 0.11$.

3.3 Discussion of Study 2

First, these results clearly support the notion that the decision to search vs. browse is strongly affected by the site information architecture in terms of

labeling and menu structure. Given broad, high-scent menus, participants searched less than 10% of the time, but they searched almost 40% of the time when faced with narrow, low-scent menus; this is a practically-significant effect.

Second, the breadth effect in this study turned out to be larger than the scent effect. While these results certainly may not generalize to all sizes of menus and all levels of semantic relatedness, it suggests that the size of the option space, and not just its scent, plays an important role in how users navigate a Web site.

4. GENERAL DISCUSSION

4.1 What Do Users Do?

On the surface, the decision whether to search or browse a Web site given some kind of target item is a complex interplay of many factors, both individual and site-related. The assumption (clearly implicit in many Web site designs) that site search is faster and/or more accurate than browsing the menus also received no empirical support in this study. The frequency and time data in Study 1 clearly suggested that the participants in this study were not irrational by often choosing to use the menus to locate items. It seems clear that they incurred no time or accuracy disadvantage by doing so.

On the individual front, some researchers (e.g., Nielsen [2000]) have suggested that roughly half of all users are search-dominant (i.e., they go straight for the search function when entering a site). However, far fewer than half of the users in Study 1 were search-dominant (see also Spool [2001]). This can be interpreted as support for the notion that particular sites play a critical role in determining searching behavior.

The notion of search-dominant users is probably misguided, as our data are consistent with hypertext research conducted in the 1980s. One hypertext study [Campagnoni and Ehrlich 1989] found that despite the use of stimuli designed to elicit both browsing and searching behavior, most users preferred to browse. Furthermore, the authors offered two explanations for the results that are consistent with the questionnaire findings from the current study. To account for the browse over search advantage, they reported that participants were often unable to formulate the query terms necessary to make effective use of the indexes (i.e., they were unclear as to the search parameters). However, for questions that resulted in uncommonly high searching behavior, they reported that the key words in the question did not appear in the tables of contents (i.e., there was a poor match of item and menu).

One researcher has classified search function deficiency under four headings: Figuring out where to search (e.g., under which product section), entering the correct words, specifying the syntax, and interpreting the results [Scanlon 2000]. Of the four problems just mentioned, knowing what words to enter appeared to be the most common problem in Study 1, but the organization and presentation of search results caused much user difficulty (and frustration) as well.

More empirical studies are necessary to fully characterize searching and browsing behavior. However, our results indicate that there is reason to doubt

several assumptions about site search functions that seem to appear on some e-commerce sites, and strongly suggest that providing good search facilities is not an acceptable substitute for taking design of the site's menus seriously.

4.2 How Do Users Decide What to Do?

Based on the results, it is clear that search rates are influenced by a multitude of factors, governed particularly by the site's information structure and layout of the home page. Our results suggest that no single factor determines whether users search or browse, but rather that multiple factors influence the decision.

Our results provided clear support for the virtues of high menu breadth. Deerskin.com and Supremevideo.com, the two most browsed sites, both possessed very broad and comprehensive menus. Apparently, participants were not bothered by the long list of menu options available to them as long as the correct choice became evident. Study 2 also yielded strong support for the virtues of high menu breadth. Menus that contained thirty options were more frequently used than those that contained only nine options. On this point, Spool (reported in Koman [1998]) has argued that "flattening" the information hierarchy by increasing menu breadth increases the probability that a user will find the content he seeks.

Although participants did consider the depth of the menus (and the associated time required to find the item) as an important factor when judging how to proceed, it is clear from their comments that information scent also strongly governed their behavior. The semantic match between the item name and menu label under which it resided on the site was a key determinant of the decision to use the menus. Clear evidence is provided when one contrasts the menu labels in Figures 6c and 6d. The substantially more informative menus of Figure 6c led to a reduced search rate relative to those of Figure 6d.

Prominence of the menu and search function were also deemed important by Study 1 participants in determining which option to choose. In accordance with predictions, the more prominent participants considered an option to be, the more likely it was used.

Finally, users' general predisposition toward using the search function was also found to be important. Participants in Study 1 often considered the search function to be less time-consuming than browsing the menus, yet many participants had serious reservations with choosing search as their product location strategy. Concerns relevant to the search parameters that should be used, the number of items that the search function would be expected to return, and the time for the search function to locate the items were all mentioned by our users. As a result, some participants were content to search on every site while others were reluctant to do so unless absolutely necessary.

4.3 Implications

Our results have several implications for the design of sites that can be both browsed and provide search functionality:

- Search behavior appears to be dependent upon characteristics of the site as well as the user.

- Searching for a product on an e-commerce site is not necessarily faster or more accurate than using the menus.
- High information scent of menu options is a key factor involved in whether a menu will be used. Participants used menus with high information scent even in the presence of a prominent search function.
- Our data suggest that very high menu breadth (i.e., greater than twenty items) does not decrease menu usage. In fact, using more specific categories may lead to better information scent and thus increased use of menus.
- There are large individual differences in users' general attitude toward using search. Assumptions about the preferences of particular user populations need to be evaluated carefully before being used as the basis for design decisions.

4.4 Recommendations

Because there is not yet a large or strong body of evidence on many of the factors suggested by these studies, proposal of design guidelines would be premature. However, it appears that increasing information scent through the selection of specific, distinctive, and semantically rich menu items leads to more effective menu usage. Furthermore, the data suggests that making the search (and browse option) clearly available to the users of a site by way of high contrast with the background and a prominent location on the page can preclude abandonment of a site if one product location method is not fruitful.

REFERENCES

BLACKMON, M. H., POLSON, P. G., KITAJIMA, M., AND LEWIS, C. 2002. Cognitive Walkthrough for the Web, In *Human Factors in Computing Systems: Proceedings of CHI 2002*. ACM, New York, 463–470.

BYRNE, M. D., JOHN, B. E., WEHRLE, N. S., AND CROW, D. C. 1999. The tangled web we wove: A taskonomy of WWW use. In *Human Factors in Computing Systems: Proceedings of CHI 99*. ACM, New York, 544–551.

CAMPAGNONI, F. R. AND EHRLICH, K. 1989. Information retrieval using a hypertext-based help system. In *Proceedings of the Twelfth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, 212–220.

CARD, S. K., PIROLLI, P., WEGE, M. V. D., MORRISON, J. B., REEDER, R. W., SCHRAEDLEY, P. K., AND BOSHART, J. 2001. Information scent as a driver of Web behavior graphs: Results of a protocol analysis method for Web usability. In *Human Factors in Computing Systems: Proceedings of CHI 2001*. ACM, New York, 498–505.

CHI, E. H., PIROLLI, P., AND PITKOW, J. 2000. The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a Web site. In *Human Factors in Computing Systems: Proceedings of CHI 2000*. ACM, New York, 161–168.

DEERWESTER, S., DUMAIS, S., FURNAS, G., LANDAUER, T., AND HARSHMAN, R. 1990. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci. 41*, 391–407.

EVANS, G. W. 1980. Environmental cognition. *Psych. Bull. 88*, 259–287.

GOLLEDGE, R. G., SMITH, T. R., PELLEGRINO, J. W., DOHERTY, S., AND MARSHALL, S. P. 1985. A conceptual model and empirical analysis of children's acquisition of spatial knowledge. *J. Exper. Psych. 5*, 125–152.

KLEINMUNTZ, D. N. AND SCHKADE, D. A. 1993. Information displays and decision processes. *Psych. Sci. 4*(4), 221–227.

KOMAN, R. 1998. The scent of information: Helping users find their way by making your site "smelly." http://webreview.com/wr/pub/98/05/15/feature/index.html.

LANDAUER, T., FOLTZ, P., AND LAHAM, D. 1998. Introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.

LARSON, K. AND CZERWINSKI, M. 1998. Web page design: Implications of memory, structure, and scent for information retrieval. In *Human Factors in Computing Systems: Proceedings of CHI 98*. ACM, New York, 25–32.

LOHSE, G. L. AND SPILLER, P. 1998. Quantifying the effect of user interface design features on cyberstore traffic and sales. In *Human Factors in Computing Systems: Proceedings of CHI 98*. ACM, New York, 211–218.

MARCHIONINI, G. 1989. Making the transition from print to electronic encyclopedias: Adaptation of mental models. *Int. J. Man-Machine Studies, 30*, 591–618.

NIELSEN, J. 2000. *Designing web usability: The practice of simplicity*. New Riders Publishing, Indianapolis, IN.

NISBETT, R. AND WILSON, T. 1977. Telling more than we can know: Verbal reports on mental processes. *Psych. Rev. 84*(3), 231–259.

NORMAN, K. L. 1991. *The psychology of menu selection: Designing cognitive control at the human/computer interface*. Ablex, Norwood, NJ.

PAYNE, J. W., BETTMAN, J. R., AND JOHNSON, E. J. 1993. *The adaptive decision maker*. Cambridge University Press, Cambridge.

PIROLLI, P. 1997. Computational models of information scent-following in a very large browsable text collection. In *Human Factors in Computing Systems: Proceedings of CHI 97*. ACM, New York, 3–10.

PIROLLI, P. AND CARD, S. 1995. Information foraging in information access environments. In *Human Factors in Computing Systems: Proceedings of CHI 95*. ACM, New York, 51–58.

PIROLLI, P. AND CARD, S. 1999. Information foraging. *Psych. Rev. 106*(4), 643–675.

POWELL, T. A. 2000. *The complete reference: Web design*. Berkeley, CA: Osborne/McGraw-Hill.

RADA, R. AND MURPHY, C. 1992. Searching versus browsing in hypertext. *Hypermedia, 4*(1), 1–30.

SANDERSON, P. M., MCNEESE, M. D., AND ZAFF, B. S. 1994. Handling complex real-world data with two cognitive engineering tools: COGENT and MacSHAPA. *Behavior Research Methods, Instruments, and Computers, 26*(2), 117–124.

SAWYER, P. 2000. Links that give off scent. User Interface Engineering, Inc. Report. North Andover, MA.

SCANLON, T. 2000. On-site searching and scent. User Interface Engineering, Inc. Report. North Andover, MA.

SCHAMBER, L., EISENBERG, M. B., AND NILAN, M. S. 1990. A re-examination of relevance: Towards a dynamic situational definition. *Information Processing and Management, 26*, 755–776.

SELLEN, A. J., MURPHY, R., AND SHAW, K. L. 2002. How Knowledge Workers Use the Web, In *Human Factors in Computing Systems: Proceedings of CHI 2002*. ACM, New York, 227–234.

SPOOL, J. 2001. Are there users who always search? http://world.std.com/~uieweb/always.htm

SPOOL, J. M. AND KLEE, M. 2000. Designing for scent. North Andover, MA: User Interface Engineering, Inc. Report.

SOTO, R. 1999. Learning and performing by exploration: Label quality measured by latent semantic analysis. In *Human Factors in Computing Systems: Proceedings of CHI 99*. ACM, New York, 418–425.

TILSON, R., DONG, J., MARTIN, S., AND KIEKE, E. 1998. A comparison of two current e-commerce sites. In *Proceedings of the ACM 16th International Conference on Systems Documentation*. ACM, New York, 87–92.

TODD, S. AND KRAMER, A. 1993. Attentional guidance in visual attention. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*. Human Factors and Ergonomics Society, Santa Monica, CA, 1378–1382.

VINSON, N. G. 1999. Design guidelines for landmarks to support navigation in virtual environments. In *Human Factors in Computing Systems: Proceedings of CHI '99*. ACM, New York, 278–285.

WANG, P. AND SOERGEL, D. 1998. A cognitive model of document use during a research project. Study I. Document selection. *J. Am. Soc. Inf. Sci. 49*(2), 115–133.

Received March 2001; revised July 2002, March 2003; accepted April 2003