INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

RICE UNIVERSITY

Banner Blindness: What Searching Users Notice and Do Not Notice on the World Wide Web

By

Jan Panero Benway

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

David M. Lane, Chair
Associate Professor of Psychology and Statistics

Daniel Osherson
Professor of Psychology

H. Albert Napier
Associate Professor of Management and Psychology

Miguel Quiñones
Associate Professor of Psychology and Management

Houston, Texas

April 1999

ABSTRACT

Banner Blindness: What Searching Users Notice and Do Not Notice on the World Wide Web

by

Jan Panero Benway

Web designers attempt to draw attention to important links by making them distinctive. However, when users are asked to find specific items, they often overlook these distinctive banners. The irony of this phenomenon I call "banner blindness" is that the user who really wants to find the information the designer has highlighted is not likely to do so. In the experiments reported here, banner blindness was investigated under controlled conditions. Banners located higher on the page and therefore farther from other links were missed more often than banners located lower on the page and closer to the other links. Banners were missed more often when located on pages containing links to categories than when located on pages with links to specific items. Users rarely noticed banners when clicking the banner was not required to accomplish a task.

Banner blindness occurred with several types of distinctive links – graphical banners that resembled advertisements, large plain-text banners and small plain-text banners that were very unlike advertisements. Increasing the perceptual grouping between the banner and the "menu" of hyperlinks helped users notice the banners only slightly more often. Adding animation to graphical banners did not help mitigate the effect. Users searching for

specific information seem to focus exclusively on the link-rich areas of the page and do not notice distinctive items outside of that area.

The last two experiments in this research focused on emphasizing one item within a menu of search-engine "hits." Three types of emphasis were used. Very large text caused a slight banner-blindness effect. Subtly large text had no effect at all. Highlighting one menu item by giving it a brightly-colored background did not cause banner blindness. In fact, it attracted the attention of users: users were more likely to select the highlighted item and did so more quickly. This type of color highlighting was most effective when it emphasized the first item in the menu. It was slightly less effective when it emphasized items in the middle of the menu.

ACKNOWLEDGMENTS

I am very grateful to a number of people that have helped with the project over the last two years. Linda Johansen was very helpful in the initial concept phase and helped me coin the phrase "banner blindness." I am also very grateful to all the people who participated in this work as subjects, to my advisor, David Lane and my other committee members. I especially appreciate my parents and my husband, Page, who have supported me throughout graduate school and life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

In a usability test of a large company's intranet[1], I discovered a very strange and, at the time, unexpected problem. I designed a simple task to test the basic design of the site. Employees could use the system to find information about training classes, and the question was, could the employees find the right links from the "home page" of the intranet to the training page? I asked each employee who participated in the usability study to "find information about classes on how to use the Internet."

The employees did not have any trouble finding their way from the home page to the training page; their problems began after that point. The top-level page on corporate training had the general design as shown in Figure 1.

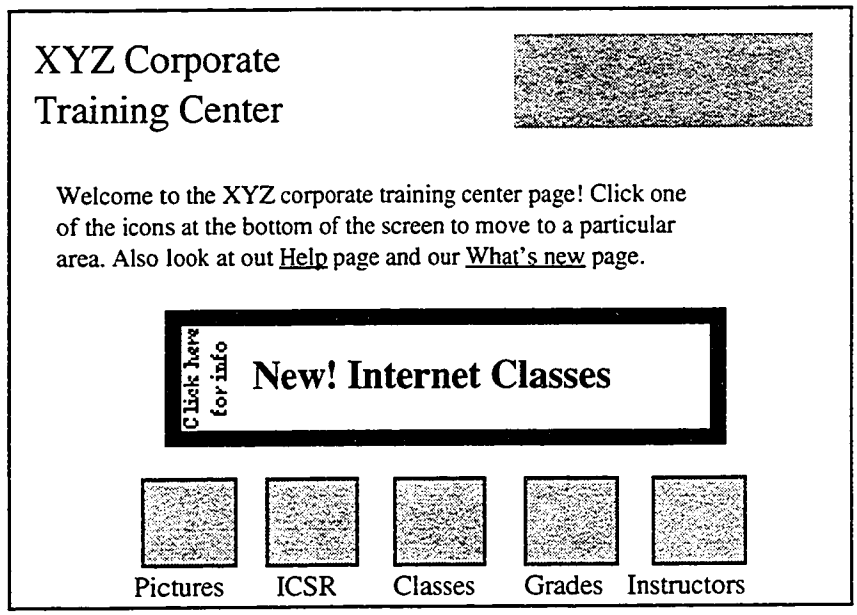


Figure 1: Home page for the training center in a corporate intranet. The box surrounding the words "New! Internet Classes" was red and was significantly brighter than the rest of the page.

[1]Intranet: A World Wide Web-based information system that is used internally in a corporation.

None of the over twenty participants in the usability study found the information on Internet classes. Each participant, in turn, clicked the "classes" link at the bottom of the training page. For some reason, Internet classes were not listed on the resultant page, and so the users never found the requested information. Clearly, the designers needed to include the information on Internet classes under the "classes" link as well as directly from the page shown in Figure 1. If they had, all of the participants would have found the information. The interesting question is why all of the participants failed to see the very prominent link to Internet classes on the training page. The designers of this page wanted to draw attention to their new Internet classes but they did not succeed.

Users of the World Wide Web do not read every word on the pages they view. Most pages are a means to an end -- a way to find information that is buried under layers of hypertext links. Jakob Nielsen (1997) has noted through usability tests of web pages that "[u]sers recklessly skip over any text that they deem to be fluff (e.g., welcome messages or introductory paragraphs) and scan for highlighted terms (e.g., hypertext links)" (paragraph 5). Others have found behavior similar to the "Internet classes" problem. An early web usability study showed that users consistently missed buttons (which looked like headers) at the top of the screen. This happened even though those same users selected another button that was literally touching the button they were looking for (Nielsen, 1994). In another study, the answer to a participant's question could be found in an animated advertising banner on the page the participant was viewing, but she never saw the banner (Spool, Scanlon, Schroeder, Snyder & DeAngelo, 1997).

Even though the advertiser or page designer wants to make a particular link extremely obvious, pulling it out by itself and putting it higher on the

page, making it brightly colored or even making it move, the user skims right by it without seeing it. I call this phenomenon "banner blindness." Banner blindness is an ironic occurrence in web interaction -- the user happens to be looking for the link that the designer *especially* wants the user to see, but that link is the one most likely to be missed.

1.1. Use of Banners in World Wide Web Page Design

1.1.1. Highlighting "News of the Day"

Many web sites present their users with a main menu of choices that are static in page location and often in content. Then, the page designers highlight a particularly important item, or oft-searched-for item by placing it apart from the menu, at the top of the page. For example, Figure 2 shows the top-level page of Yahoo! (Yahoo! Inc., 1997) on a day in late September. During this time Major League Baseball is important and interesting news. Therefore, Yahoo! has provided a quick and "obvious" link to baseball news right at the top of the page. Although I have not tested this page for usability, the results described earlier in this paper suggest that Yahoo! users who are explicitly looking for baseball news may skip immediately down to the menu below and not realize that Yahoo! had provided a quick and easy link for them at the top.

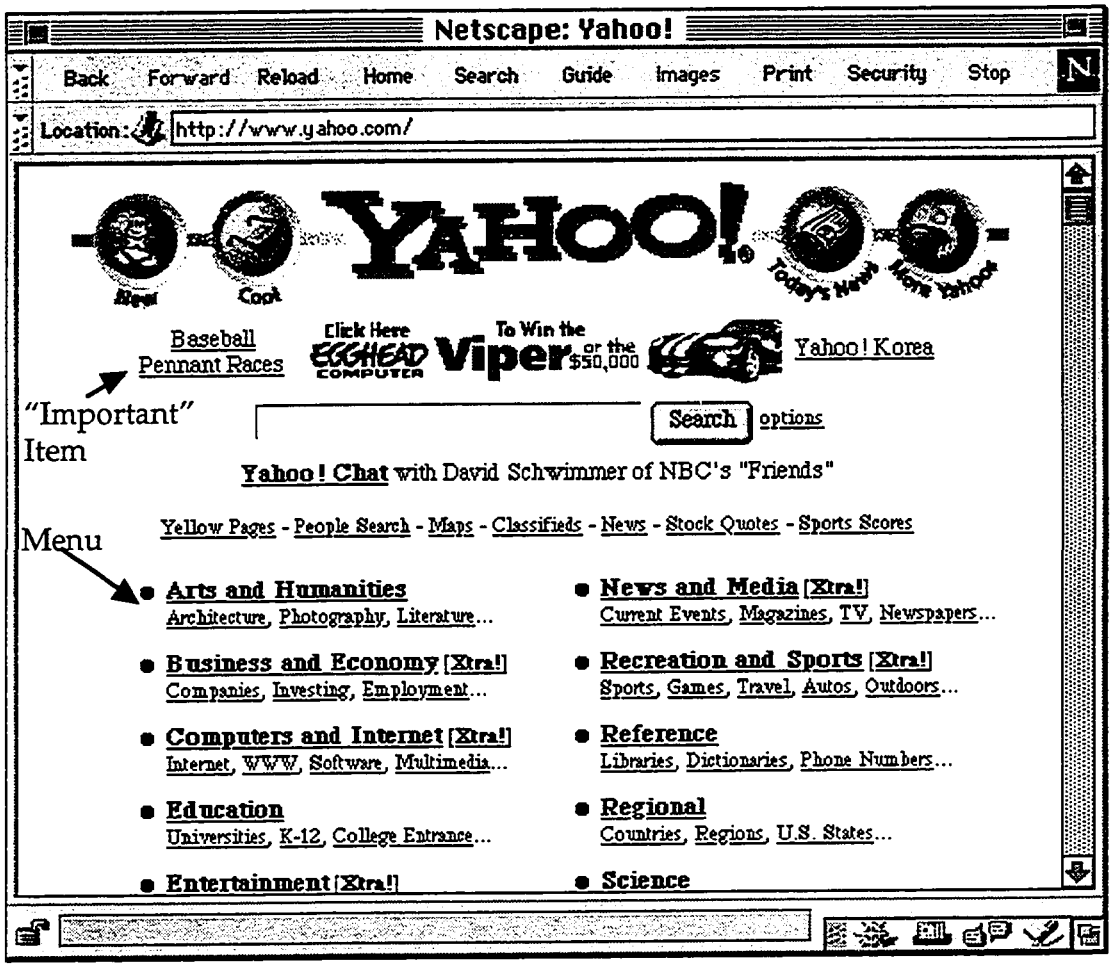


Figure 2: On a day where baseball was at the "top of the news," the Yahoo web site has provided a direct link to baseball, separate from the main menu on the page (Yahoo! Inc., 1997).

As this example shows, I am using the term "banner" broadly - not just to refer to rectangular advertising banners, but to any item meant to stand out like this small, textual baseball "banner." This is not to neglect how the banner blindness phenomenon might apply to advertising, but it is important to note that its application is wider than advertising alone.

Another way web designers try to attract attention to important items is with a large graphic. Figure 3 shows the home page for Salon Magazine, a general interest on-line magazine. Salon uses a graphic at the top of the page to advertise and link to its featured article of the day. The other articles are

listed in a menu lower down on the left side of the screen. This graphical link can be considered a "banner" because it is visually separated from the main menu and is meant to attract attention to the most important article of the day.

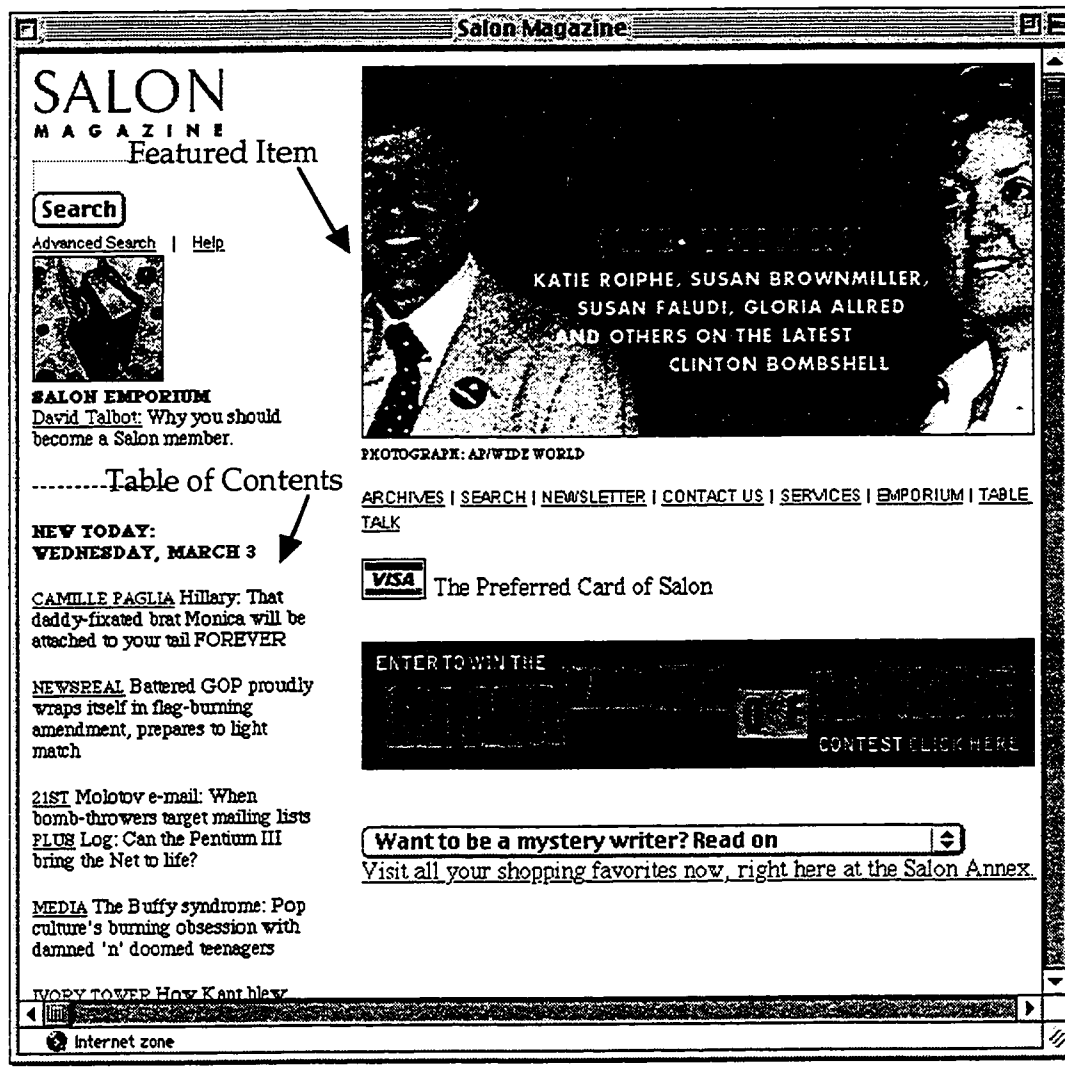


Figure 3: Salon Magazine uses a large graphic to link to its featured item of the day. The other articles are listed separately, on the left side of the page. (Salon Internet Inc., 1999).

The corporate training, Yahoo, and Salon home pages all follow conventional web guidelines which usually recommend that to make an important item stand out, it should be near the top, and be large and/or brightly colored. For example, the Ameritech web design guidelines state:

> In general, the larger an item is, the greater its perceived visual importance and likelihood of attracting attention. Make sure that items of greatest importance are easy to see, and clearly distinguished from other items. (Detweiler & Omanson, 1996)

Experiment 1 investigated the validity of this advice.

2. EXPERIMENT 1

The first goal of this research was to determine whether banner blindness could be recreated in a laboratory setting. This experiment focused on two types of banners: rectangular graphic banners like the ones that intranet usability participants failed to see (Figure 1) and graphical features like those used in Salon Magazine (Figure 3). Non-graphical banners were studied in Experiment 2.

Experienced participants were used because most of the participants in the usability test had been novices, and I wanted to ensure that the phenomenon was not an artifact that disappeared quickly as a user learned more about the web.

2.1. Method

Six subjects, all with moderate to extensive web experience, volunteered to participate. All were working professionals who ranged in age from the late 20's to the late 30's. They interacted with a web site designed for this study. The web site was strictly hierarchical, and contained three levels in addition to the bottom-level "content" information. A schematic diagram of the site's organization is shown in Figure 4.

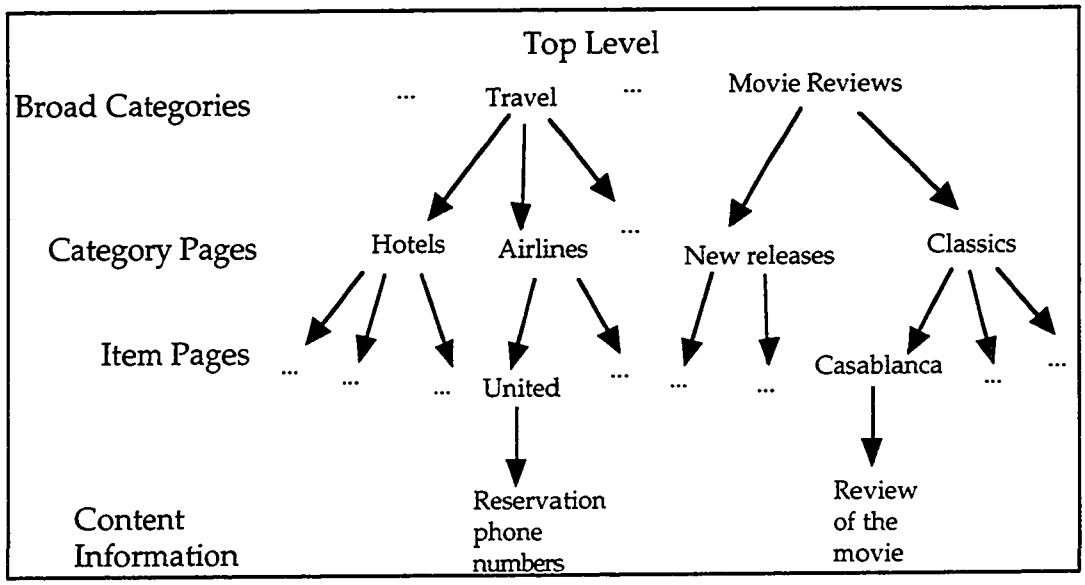


Figure 4: Organization of the web site used in Experiment 1.

Banners were placed throughout the site. Some of the banners were large rectangles with a red background, as shown in Figure 5. Other banners were "graphic" banners as shown in Figure 6. Both types of banners appeared either as category-level pages of the site or on the item-level pages (see Figure 4). Red rectangular banners appeared either at the very top of the page or immediately above the menu. Graphical banners appeared to the left of the menu.

Thus, the banners varied over two variables. The "location" variable had three levels: red banners at the top of the page (high), red banners in the middle of the page (low), and graphical banners to the left of the menu (graphic). The "hierarchy" variable had two levels: banners on item- and category-level pages. Figure 5 illustrates a low banner on a category-level page, and Figure 6 illustrates a graphic banner on an item-level page.

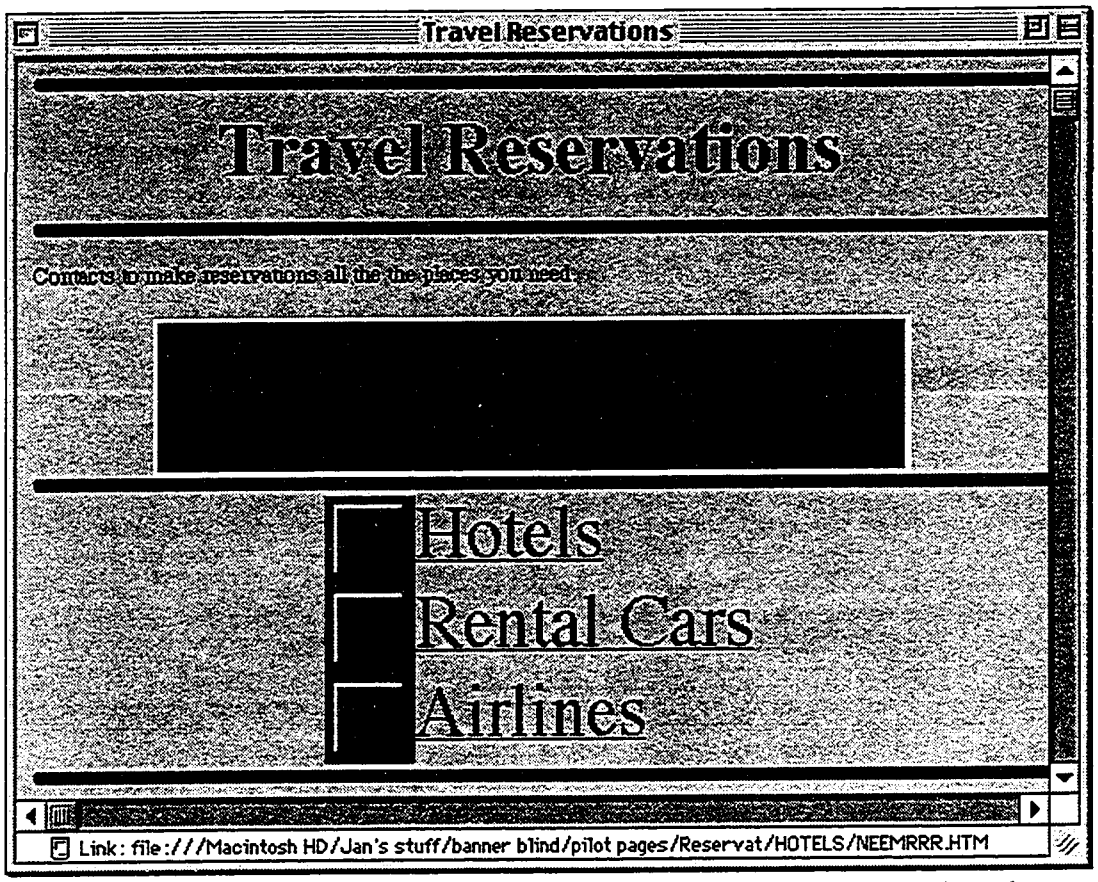


Figure 5: A page from the web site used in Experiment 1. The search task was to "Find the email address for Neemrana Hotels." The banner had a bright red background.

Figure 6: A page from the web site used in Experiment 1. The search task was to "Find information about the HCI International Conference."

There were banner trials, control trials, and missing trials. Banner trials were the experimental trials. In banner trials, the appropriate banner had to be clicked in order for the subject to find the requested information. In control trials, the required information could be found by clicking text menu choices only. Banners were sometimes present, but were never needed to find the information. In missing trials, subjects were asked to find information that was not available on the web site. The subjects were warned that in some cases the information they were asked to find would not be available.

The subjects were asked to perform 24 searches for specific pieces of information. The searches were given in random order, with the constraint that the first task be a control trial and the second task be a "missing" trial. There were six banner trials, one of each type according to the 3 (location) x 2 (hierarchy) design. There were 14 control trials and four missing trials.

2.2. Results

There was substantial banner blindness. The subjects located 93.8% of the control targets, and fewer of the banner items. Table 1 shows the success rate for each banner type. The banners on the category-page level were most missed, and banners higher on the page were missed more than the banners immediately above the menus. Graphics to the left of the menu were missed least often. Overall, the success rate for controls was significantly greater than for the banner items, $t(5) = 2.64$, $p = 0.046$.

Table 1: Rate at which subjects located the banner targets. Success on controls was 93.8%.

	Hierarchy level		
Location	Category	Item	Means
High	33.3%	66.7%	50.0%
Low	50.0%	83.3%	66.7%
Graphic	50.0%	100%	75.0%
Means	44.4%	83.3%	63.9%

Two secondary measures were taken to determine whether the subjects had difficulty in locating the banners they did find. After they found each item, they were asked to rate how difficult the item was to find. The subjective ratings were made on a five-point Likert scale where 1 indicated "very easy to find" and 5 indicted "very difficult to find." No ratings were made when the subject failed to find the item. Also, the tasks were timed to

determine whether subjects spent more time looking for banners than for the control objects.

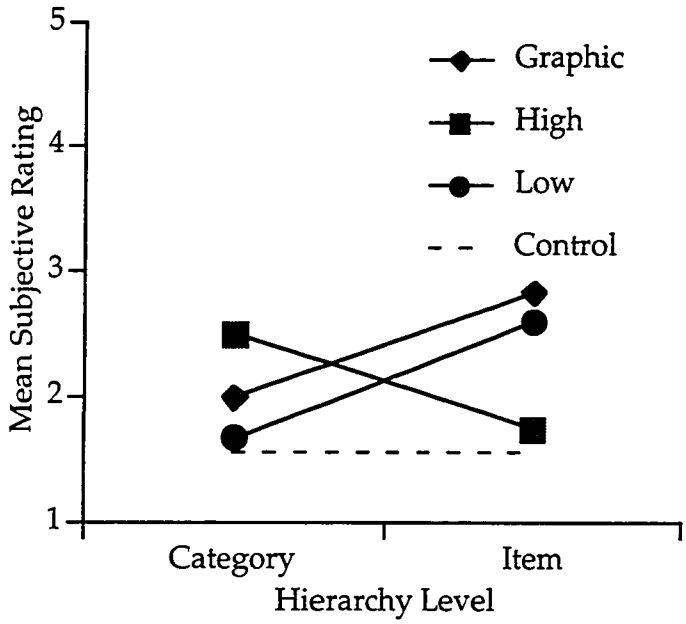


Figure 7: Mean rating of how difficult it was to find the items, successful finds only.

Figure 7 shows how each type of banner was rated. The control items were always rated easier to find than the banner items, although in some cases only marginally so. The overall difference was significant, $t(5) = 2.81$, $p = 0.037$. There appears to be an interaction between the hierarchy level and the distance from the menu. When the banner was found low on the page whether it was a rectangular banner above the menu (low) or a graphic to the left of the menu (graphic), it was rated easier to find on the category level than on the item level. Conversely, when the banner was found high on the page (far from the menu), it was faster to find on the item level. It is difficult to tell whether this interaction is reliable. Because most subjects failed to find several items, the amount of missing data makes a significance test infeasible. It is interesting to note, though, that these results differ in direction from the success data (Table 1) in which there was no interaction.

"Scan time" is the time the subjects spent looking at the page containing the banner before they clicked it. This accounts for only the last time they viewed the page during a given task. The mean scan time for control trials was 3.32 seconds. This was faster than the banner trials which averaged 7.88 seconds. This difference was significant, $t(5) = 4.56$, $p = 0.008$. As shown in Figure 8, scan time for category-level banners was fast, even marginally faster than controls. The scan time for item-level pages, however, was considerably longer. Once again, the direction of effects here are different from the success data.

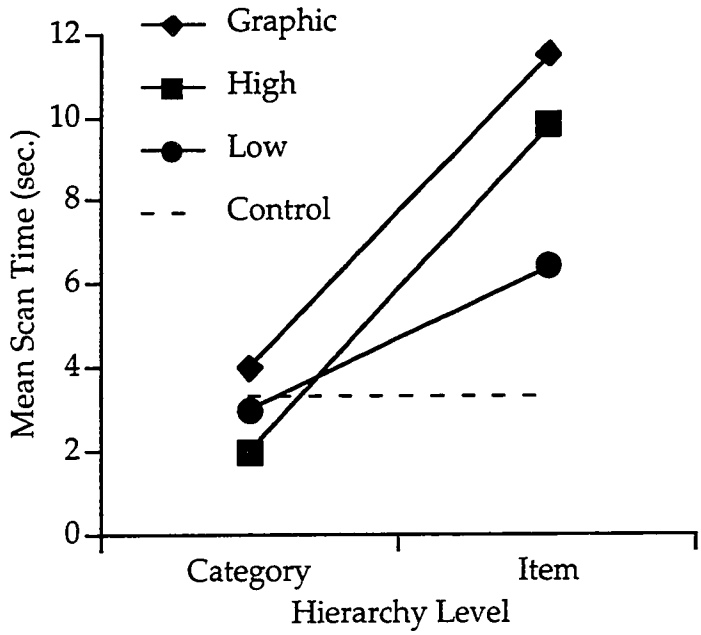


Figure 8: Mean scan time to locate the items, successful finds only.

As suggested above, the degree of success subjects had with each type of banner correlated positively with the difficulty ratings those items received (r = 0.51). Also, success correlated positively with scan time, r = 0.87. This seems to imply that the subjects who succeeded more often on the banner trials may simply have been trying harder. However, when comparing more and less successful *subjects*, it seems that those people who found more of the banner items tended to give lower difficulty ratings (r = -0.80) and were faster in

making selections ($r = -0.78$). Therefore, it seems that there were differences among subjects in how much banner blindness they experienced. Those who were most likely to fail on the banner tasks also tended to take longer and rate more difficult the banner tasks they did succeed on. The initial contradiction wherein those items that were found most often also tended to take longer and be rated difficult is probably a result of more banner-blind-susceptible subjects being included in that data.

2.3. Discussion

The banners used in this experiment seem like they should have been obvious to users: they were large, brightly colored, and stood out from everything else on the page. However, this study shows that they were frequently missed by the users who were specifically looking for the information they contained. Not only did the banners fail to jump out at the users and grab attention, they were frequently ignored. Even in cases where the participants located the necessary banners, they rated those tasks as being more difficult than the controls, and it took longer to find them.

This first experiment examined only a limited range of salient items, namely large graphics and red rectangular banners. There are several possible explanations for why the banners were ignored. For instance, many of the banners in this study resembled advertisements, and it is possible that the participants in this study have learned to ignore advertisements when searching for information on the web. A second possibility is that since searchers tend to look for linked text, they focus on small, blue text objects. Perhaps banners that more greatly resemble linked text would be less likely to be ignored. A third possibility is based on perceptual grouping. Although some design guidelines recommend that important items be set apart from other objects, perhaps a better strategy would be to increase the perceptual

grouping between the menu of links, which searching users presumably focus on, and the salient banner.

2.3.1. Web Advertising: A Special Case of Banner Blindness

Web advertising is one of the most common uses of banners on web pages. The advertiser pays the owner of the page that displays the banner, and hopes that web users will click the banner and visit the advertiser's web page for more information.

Only a small percentage of users who view advertising banners click them (Briggs & Hollis, 1997). Exact measures of the typical of users who typically click banner ads (referred to as "click-through) are difficult to find, but at least one advertising company guarantees its corporate customers a minimum click-though of 1% (DoubleClick Inc., 1997). On the other hand, at least one study has shown that web users who do not click advertising banners are influenced by them. Unclicked advertising banners cause increased brand awareness and purchasing probability (Briggs & Hollis, 1997).

Certain aspects of banner design, such as bright coloring and animation, are intended to attract attention to a banner and increase its click-through rate. Other aspects of banner design such as posing a question or using a phrase like "click here" also increase click-through rates (I/PRO Research/DoubleClick, 1996). The fact that semantic changes to ads increase their click-through rates suggests that web users do see and process these banners, even if they do not always select them.

Why then do usability studies show that web users do not see such banners? Usability test participants are given specific targets to search for and they conduct goal-directed scans of web pages looking for that topic. Web users in general are sometimes goal directed and sometimes "browsing" or just looking for something interesting. A browsing user is more likely to be

willing to diverge from a path to follow banner links that seem appealing. Browsing users should be more likely to click banners, but that does not necessarily explain why usability test participants don't seem to even *see* advertising banners.

It could be that experienced web users learn to block out advertising banners, knowing that they are rarely helpful. Banners that move may be even more likely to be blocked out by goal-directed searchers because the movement is distracting (Spool et al., 1997). But the participants in the usability study described at the beginning of this paper were using an advertisement-free intranet, and many of the participants were novice web users, who, because they had never used the web before would not have learned to avoid advertising banners. Therefore, there is reason to believe that banner blindness is not purely an advertising effect. It should apply to any "distinctive" link which is physically and perceptually separated from the main menu on a page.

It could be that advertisers do not care whether they attract goal-directed users. Most people browse from time to time, and so attracting those users to the advertiser's site may be enough. However, web advertisers do try to attract goal-directed users. Some advertisers target their banner ads toward people who search for particular terms, as shown in Figure 9. In this case, the user searched AltaVista for the term "weddings" and, in addition to AltaVista's search results, the page returned an advertisement for a wedding-oriented web site (Compaq Computer Corporation, 1999).

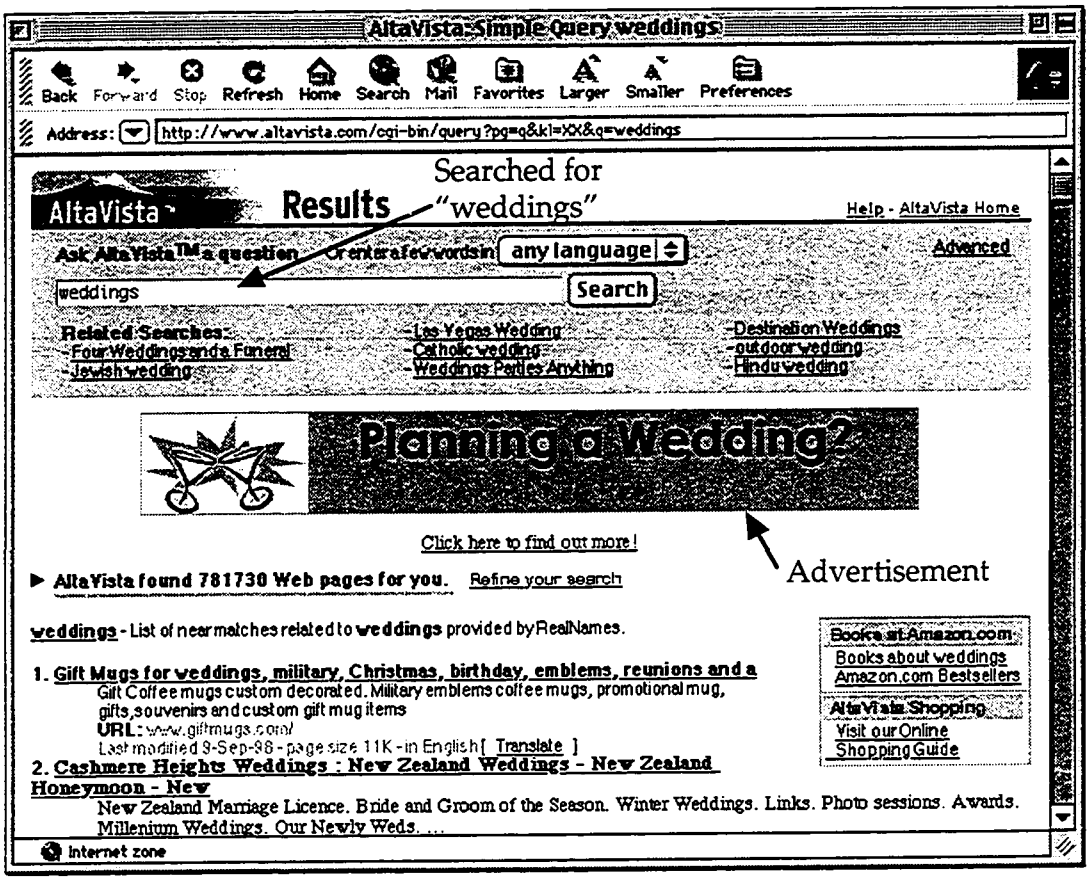


Figure 9: The user searched AltaVista for the keyword "weddings" and was presented with a targeted banner advertisement in addition to the results of the search (Compaq Computer Corporation, 1999).

The owner of the wedding-oriented page advertised on the AltaVista site hopes users who want wedding information will choose the most salient item that resulted from the search, the banner advertisement. The question is whether making the link to that site salient and distinctive in this way will in fact attract the searcher's attention.

Experiment 2 examined the relationship between advertising and banner blindness in two ways. The first way was to determine to what extent banner blindness is a result of web users learning to ignore advertising. If users fail to notice salient text on a web page even though it looks nothing like and advertisement, then ignoring banners is probably not just a result of wanting to avoid advertising. Experiment 2 also examined the effect of

advertising by looking at whether changes in traditional advertisement banners influence the likelihood that searching users will notice them. For example, will searching users be more or less likely to remember and advertisement that contains animation? Does changing the perceptual grouping on a page influence whether an advertising or non-advertising text banner is noticed?

Perceptual grouping may be a key concept in the banner blindness phenomenon. Design guidelines recommend that placing an item in its own perceptual group makes it stand out and seem important. The results of Experiment 1 may indicate that that recommendation is incorrect, but an empirical comparison of different grouping arrangements is needed to determine the role of perceptual grouping in banner blindness.

2.3.2. Perceptual Grouping

Distinctive items become distinct because the other items form perceptual groups which exclude the distinctive item (Duncan & Humphreys, 1989). Banks and Prinzmetal (1976) showed that when the distracter set forms a perceptual group, the target is easier to identify. One way to make an item distinct from the others is to physically separate it. The item of interest is thus in a "group" by itself (making it distinctive), and the remaining items are grouped by proximity. This could attract more attention to the distinctive item, as it did in the study by Banks and Prinzmetal, or it could have indeterminate effects just as other dimensions that cause distinctiveness do. However, the studies on the effectiveness of distinctiveness have largely ignored the dimension of spacing.

Extensive basic research in attention has shown grouping of elements to be a basic function. Because attention is a limited resource, it cannot be simultaneously be focused on the whole visual field. Before focal attention is

allocated, preattentive processes (Neisser, 1967) break the field into figural units. A review of research on visual search shows that virtually all models of visual search agree that there is a series of stages of attention and that the first, preattentive stage is parallel. This preattentive stage guides attention, restricting it to the part of the visual field most likely to contain the target (Wolfe, 1994).

The more a particular grouping property (color, orientation, proximity) helps form a group, the more likely that group is to attract attention when the viewer has a peripheral view of it (Beck, 1972). Once the groups are identified, the search process proceeds serially, as the viewer attends to one group after another (Treisman, 1982). Only after the viewer selects a group and focuses attention on it does he or she processes the items within the group, either serially or in parallel (Kahneman & Henik, 1977).

In computer displays, most discussion of perceptual grouping uses proximity as the display dimension which groups the items. A major issue in grouping by proximity is the amount of proximity required in order for objects to be perceived as a group. Based on previous research (Cropper & Evans, 1968; Woodward, 1972) and his own empirical results, Tullis (1983; 1988) concluded that a proximal group should be within a 5 degree diameter circle around the point of fixation. This is because visual acuity is cut in half at about 2.5 degrees from the point of fixation. Assuming a typical viewing distance of about 475 mm, a group of objects within a circle of 41.8 mm can be considered to be "near" one another (Tullis, 1983). However, Treisman (1982) found that varying the distance between different groups did not affect the seriality of the search between different groups. Groups were searched serially rather than in parallel whether or not eye movements were required to examine the different groups.

Tullis concluded that groups should fit within 5 degrees of viewing angle and be surrounded by white space. He does not indicate how much white space is required. However, his studies were performed in a VDT display in which all characters were the same size. A single space all the way around a group seems to delimit the groups he shows in his examples. Other screen design researchers have also designated that groups should be surrounded by "white space" (Streveler & Wasserman, 1984) or have given specific recommendations for numbers of spaces needed on a VDT display without much empirical justification (Pulat & Nwankwo, 1987). A study comparing an expert system to human judgment found that humans use more than the existence of white space to define a group; they also consider hierarchical and other spatial variables to mark where groups exist (Williams & Leaf, 1995). Similar results were found by Bednall (1992), who showed that subjects were equally aided by groups that were created through white space between the groups and by groups that were created by hierarchical organization within the groups. Therefore it appears that grouping may be more complex than the simple addition of white space.

Grouping is not a binary judgment. Studies have shown that the amount of grouping between objects decreases as the items are moved further apart (Gillam, 1972; Pomerantz & Schwaitzberg, 1975). The exact amount of space that can exist before grouping breaks down is dependent on the stimuli. Gillam (1981) showed that for simple stimuli, grouping is a function of the ratio of the size of the objects and the spacing between them. Thus, larger objects can group across more distance than can smaller objects. Shipley and Kellman (1992) showed that the ratio rule extends to more complex illusory stimuli. However, a magic grouping ratio does not exist. The amount of space that items can group across is dependent on the particular stimuli. King (1990)

showed that items that create a "good gestalt" can group over larger distances. For instance, these two items: [ ], form a relatively good gestalt and thus can be seen to group over a larger distance than can these two items: ΔΔ. There is thus no constant grouping ratio because these two sets of stimuli are of similar size but group over different amounts of white space.

Because of the difficulty in defining exactly what a perceptual group is, grouping could be acting on computer interfaces in difficult-to-determine ways. One possible explanation for banner blindness is that the distinctiveness of the banner is actually working against the important item because searchers are avoiding distinctive item to focus on link-rich areas of the page. Or, the difficulty could be that the although the banner seems to be distinctive through color and placement, it actually is grouped together with less important information. Studies of web users show that when they read web pages for information, they tend to skip over anything deemed to be "fluff" (Morkes & Nielsen, 1997; Nielsen, 1997). This can include titles, headers, introductory information or marketing material. According to the visual search model of Duncan and Humphreys (1989), when people are first exposed to a visual scene they form a "perceptual description" in which the scene is segmented into structural units. These units are strongly influenced by perceptual groups, particularly proximal groups. Weights are then assigned to each perceptual group based on the probability that they contain the target. Groups tend to be accepted or rejected as a whole. A process of spreading inhibition dictates that if a group is given a low weight, all the elements in the group are likely to be allocated low levels of attention. Therefore, if a banner on a web page forms a perceptual group with other information (titles, marketing hyperbole), the entire group, including the banner, is likely to be ignored. Figure 10 shows two possible segmentations of the "Corporate

Training Center" web page described at the beginning of this paper. In arrangement A, the banner is in a group by itself, and thus it is distinctive. In arrangement B, the banner is in the same group with some introductory material, which many web users may judge to be "fluff" and will ignore it. If arrangement B prevails, the inhibition of the introductory material may spread to the banner, making it unlikely that that area of the screen will receive focal attention. If arrangement A prevails, the banner still may be ignored, but it has a somewhat lower probability of being inhibited.

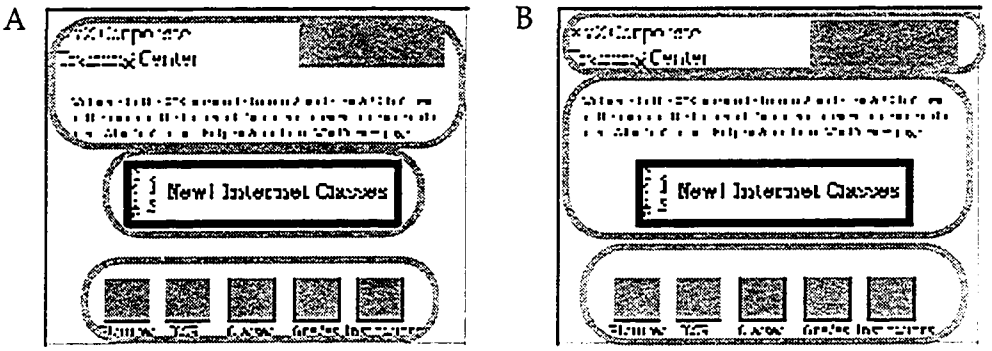


Figure 10: Thumbnail sketches of the schematic design of a web page shown in Figure 1, with some possible perceptual groups circled in "——."

Experiment 2 examines a way to increase the amount of perceptual grouping between and banner and the menu, and tests the effect of changing the grouping on nonadvertising and advertising banners.

3. EXPERIMENT 2

3.1. Method

Seventy two undergraduate psychology students from Rice University participated in exchange for course credit. 55% of the subjects were female. All the subjects had previous web experience, and 90% of them used the web every day or at least several times a week.

This experiment had two main parts, a set of "non-ad" trials and a set of "ad" trials. The subjects participated in both parts.

3.1.1. Materials

The subjects' task for each trial was to find a specific piece of information using the web site provided for them. They navigated through three levels of category choices in each trial. The first page contained three categories to choose from, the second contained 18 categories, and the third level contained five choices. The second level is the main level of interest because the banners were placed there.

3.1.1.1. Non-Ad Trials

In the non-ad trials, a banner placed on the second-level page could help the subject find the correct category. An example is shown in Figure 11. The task in this case was to find information on the book *Coaching Youth Basketball*. The subject has already selected the "books" category, and now has to select the "sports" category from the eighteen choices provided. The "banner" on this page, which states "Coaching Youth Basketball in Item C5" can help the subject locate the correct category quickly. Subjects who do not notice the banner will have to examine many of the categories before locating the correct one.

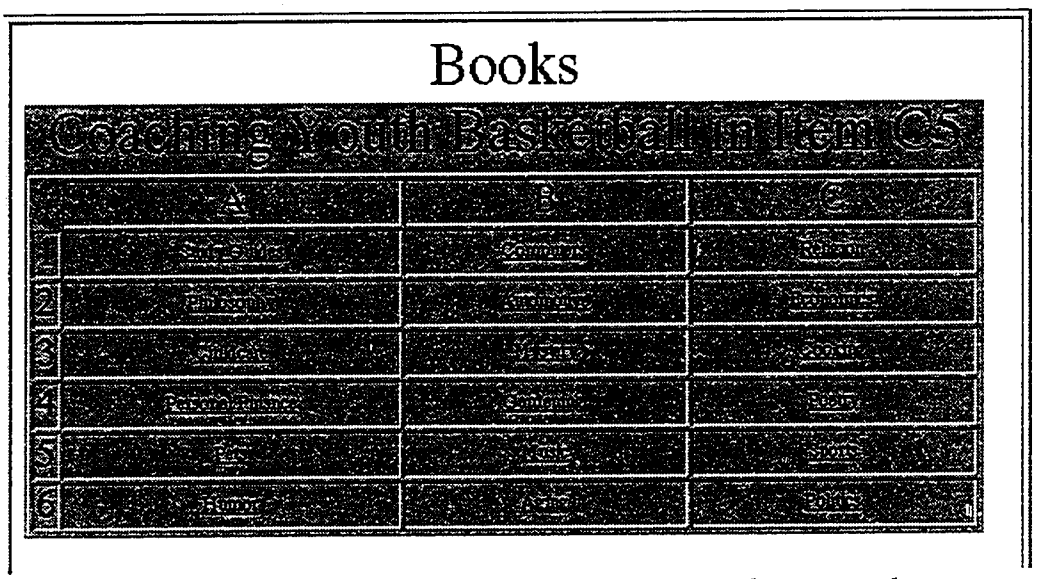


Figure 11: A page in a non-ad trial, with a large text banner and menu grouping. The underlined text was blue, the color typically used for links. The banner text was also blue. The remaining text was black.

Two between-subjects variables were manipulated in the non-ad portion of the experiment. The banner shown in Figure 11 is in the "large text" format. Other subjects saw banners instead in either the small text format, as shown in Figure 12, or the graphical format as shown in Figure 13. The second variable manipulated was perceptual grouping. The gestalt

principle of "common region" (Palmer, 1992) was used to create grouping between the banner and the menu of choices (menu grouping) or between the banner and the title (title grouping). Menu grouping is shown in Figure 11, and title grouping is shown in Figure 12.

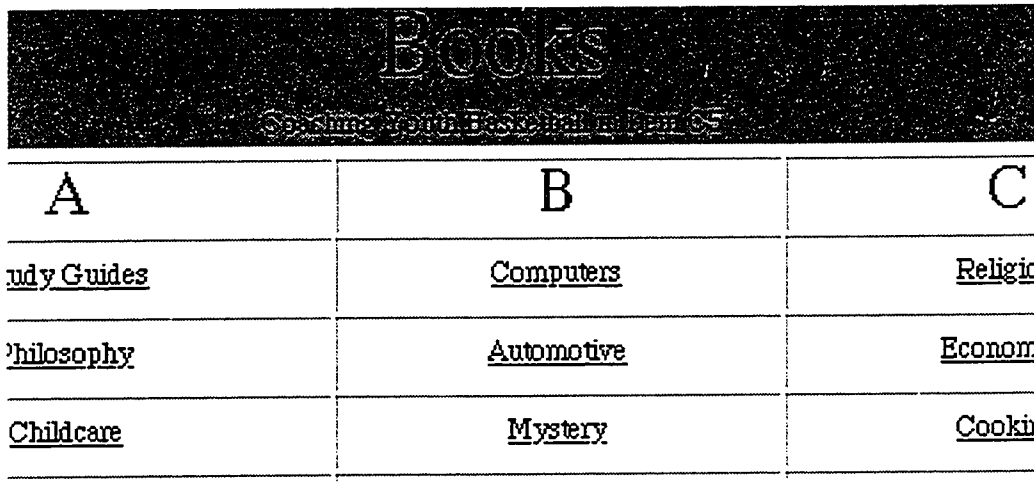


Figure 12: A small text banner with title grouping. The links and banner were blue.

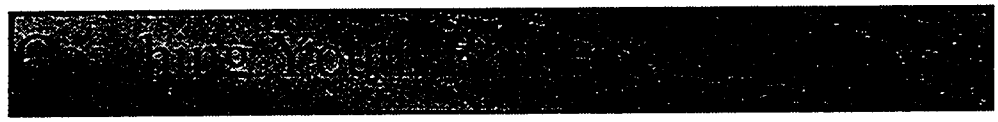


Figure 13: A graphical non-ad banner. The text of the banner was black and most of the background was blue, with a yellow streak.

3.1.1.2. Advertising Trials

In the advertising trials, the tasks were similar to the non-advertising trials. The difference was that the banners that the subjects were exposed to were always irrelevant to the search task. An example is shown in Figure 14. In this case, the subject is looking for a book titled *How to Prepare for the SAT I*, and while searching for the correct category, is exposed to an advertising banner for "Jimmy Dean Tastefuls."

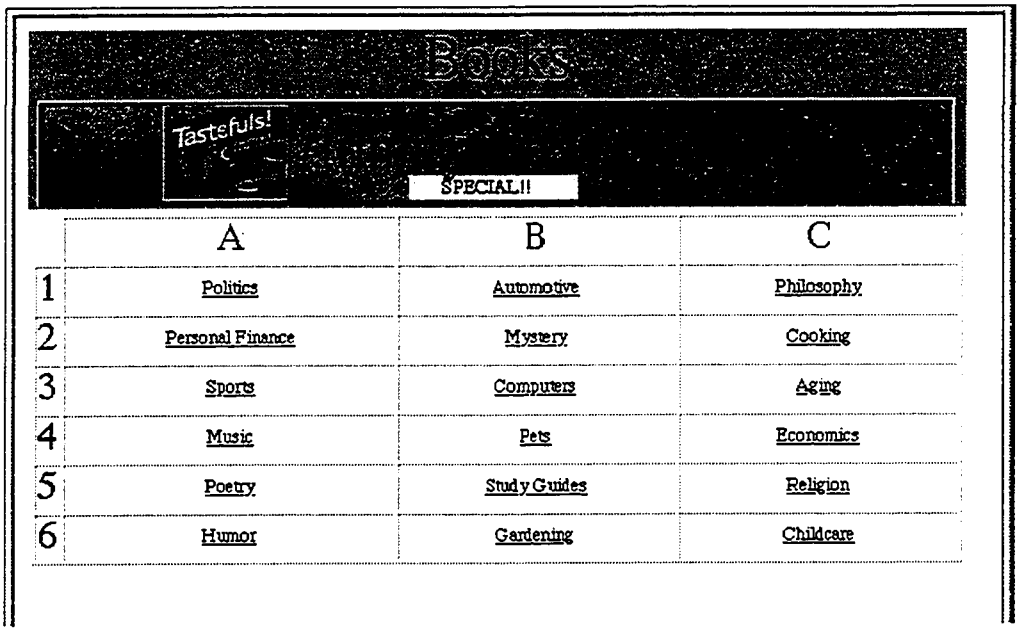


Figure 14: A page in a advertising trial, using title grouping. The box surrounding "Special!!" was yellow.

Three within-subjects variables were manipulated in the advertising trials. All the banners presented were in a graphical style, as shown in Figure 14, but some were still and some were animated. For the animated banners, the word "Special!!" scrolled across the small yellow box at the bottom of the banner.

A second variable was perceptual grouping, which was created using common region, just as in the non-ad trials. Half of the advertising trials used title grouping, and half used menu grouping.

The third variable was manipulated in the recognition memory test that occurred after the search trials were completed. In some of the recognition questions, the subjects were shown the same graphical representation of the banner shown during the presentation (*high context*). In other recognition questions, the subjects were given the content of the ad without the graphical context (*low context*). The recognition test is described in greater detail below.

3.1.2. Procedure

Before the subjects began the search trials, they were given pre-training in using the grid layout that was used in the critical eighteen-item category pages. Because the non-ad banners referred to specific categories by grid location, it was important that the subjects understand what was meant by "Item C5," for example. The subjects were shown a page containing a grid like the one shown in Figure 14, and the computer directed them to click the items A6, C3, and then B1 in order to "familiarize [themselves] with the web pages." If the subjects chose the wrong category or took too long to make a selection, they were given further help in understanding the directions.

After the practice session, the subjects began the search trials. The search tasks were numbered and printed on index cards. For each trial, the subject would read the task (e.g., "Find information about Curry Travel Agency") and then click a "begin" button on the computer screen. A screen containing three hyperlinks would appear, giving the choice between "Books", "Travel", or "Household Items." After the subject selected the correct category, a page containing 18 categories as shown in Figure 14 appeared. After the correct category was selected on that page, the subject was required to make a selection from five specific items, including the one named in the task (e.g., "Curry Travel Agency"). When the correct specific

item was selected, the subject was shown a web page containing information on the appropriate topic. The subjects were instructed just to look at this page and not to do anything. After a short period of time, the page disappeared automatically and the subject was brought to a page introducing the next task.

On all the pages requiring the subjects to make a selection, only one selection was valid. If subjects picked incorrect items, they were brought to a page that said they had made an incorrect selection, and they were asked to go back and try again.

The experiment began with three "buffer" trials. On these trials, no banners of any kind appeared on the second-level pages. The buffer trials were followed by three non-ad trials, 24 ad trials, and in some cases, three additional buffer trials. The order of the ad-trial and non-ad-trial blocks was counterbalanced, so that half the subjects completed the ad trials first and half completed the non-ad trials first. The search trials were followed by a recognition test related to the ad trials. In order to maintain a constant time delay between the ad trials and the test, subjects who completed the ad trials last also completed three additional buffer trials before the recognition test.

3.1.2.1 Non-Ad Trials

The three non-ad trials consisted of one "critical" experimental trial and two controls. The experimental trial used a helpful banner which could help the subject find the correct category quickly. One of the control items (unhelpful banner) directed the subject to a particular category, but the banner did not help the subject. For example, the task might be to "Find information about *Coaching Youth Basketball*" but the banner would state "*Mind Design* in Item A2." The second control item contained a banner that was completely irrelevant. It stated "Eat Pace Picante Sauce." In all the non-ad trials the correct category was located in row 5 of the grid. Within each group, all three

trials were of the same format (large text, small text, or graphical, and with menu grouping or title grouping) and the order of the three trial types was counterbalanced using a Latin Squares design. The order of the content of the tasks remained constant across subjects, meaning that the basketball search task was always first but the content of the banner in that trial varied.

3.1.2.2. Advertising Trials and Recognition Test

The advertising trial block variables were manipulated within subjects. The 2 (animation) x 2 (grouping) x 2 (test format) design created eight trial types. The order of these types was counterbalanced using a Latin Squares design, and each subject completed three blocks of eight, for a total of 24 ad trials. Forty-eight advertising banners were created, and half were randomly selected for each subject to be presented or to be distracters in the recognition test. An animated and an unanimated version of each ad existed, so that the appropriate version was placed according the trial type. The order in which the ads were presented was random for each subject. The content of the tasks was again constant across subjects. The location of the correct category on the 18-item page was random for the ad trials, and it could occur in any of the possible locations.

After all the search trials were completed, the subjects were given a surprise recognition test, administered automatically by the computer. All 48 advertisements were presented to the subjects, in random order. The testing format varied. For the high-context items, the banners were shown exactly as they had been presented earlier, complete with the presence or absence of animation. An example of a high-context item is shown in Figure 15.

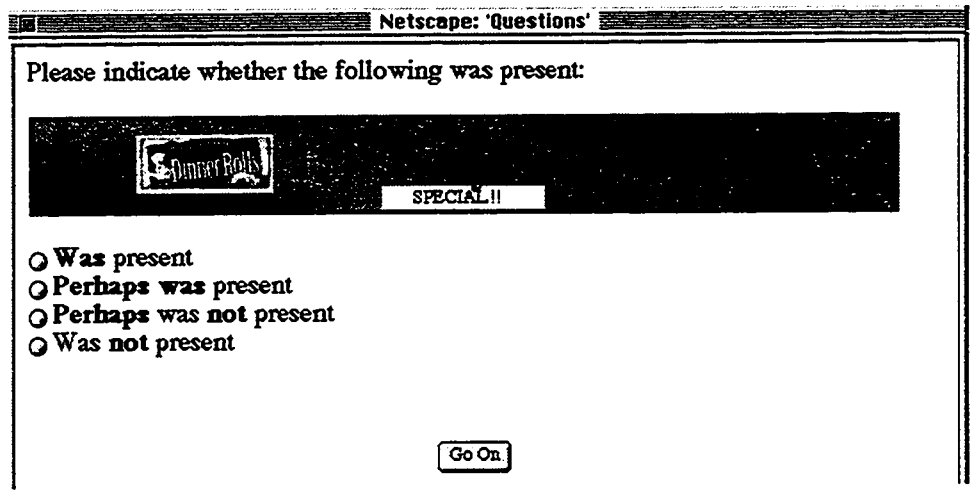


Figure 15: A high-context item in the recognition test.

The low-context recognition test items had two parts. The first part asked the subjects to rate whether they believed they had been presented a brand name (Figure 16). Then, they were forced to make a choice between two possible product names (Figure 17). The purpose of the low context items was to probe the subjects for slightly deeper processing of the ads, at two levels of detail (brand name and product name).

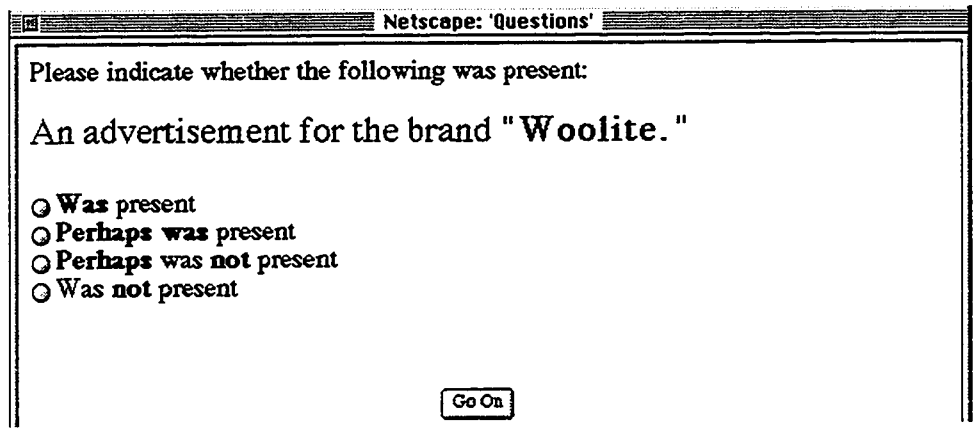


Figure 16: A low-context recognition test item, part 1.

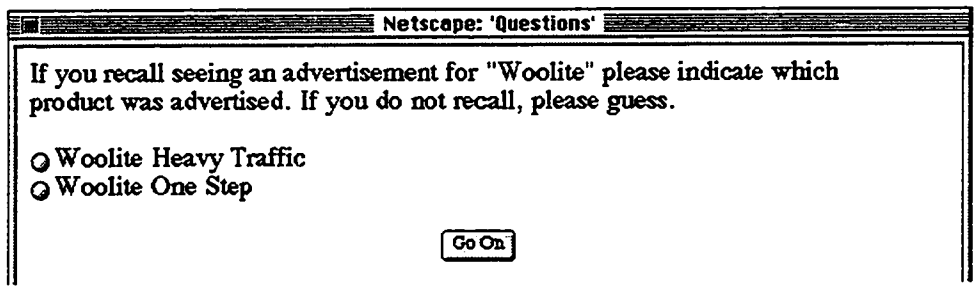


Figure 17: A low-context recognition test item, part 2.

For each ad, only one product name was used, but the two products were pre-tested to be approximately equal in familiarity to Rice students. Subjects were asked to randomly choose between the two brand names if they didn't recall the ad. This way, a base rate for which product would be more likely to be selected could be established for the distracter ads. The order of the two product names was randomly determined for each test item.

The distracter ads were evenly split so that an equal proportion of them were displayed with and without animation, and in high and low context.

3.1.2.3. Questionnaire

After the recognition test, the subjects filled out an on-line questionnaire that asked for demographic information, web experience, opinions about web features (like animation), and some specific questions about the experiment. Subjects were also given the opportunity to type comments about the experiment. The questionnaire and a summary of questionnaire results are shown in Appendix A.

3.2 Results

3.2.1. Non-Advertising Trials

Subjects in all six conditions experienced pronounced banner blindness. Figure 18 shows the difference scores for each of the groups. Difference scores were computed for each subject by subtracting the

experimental trial time from the average of the two control (unhelpful and irrelevant) trial times.

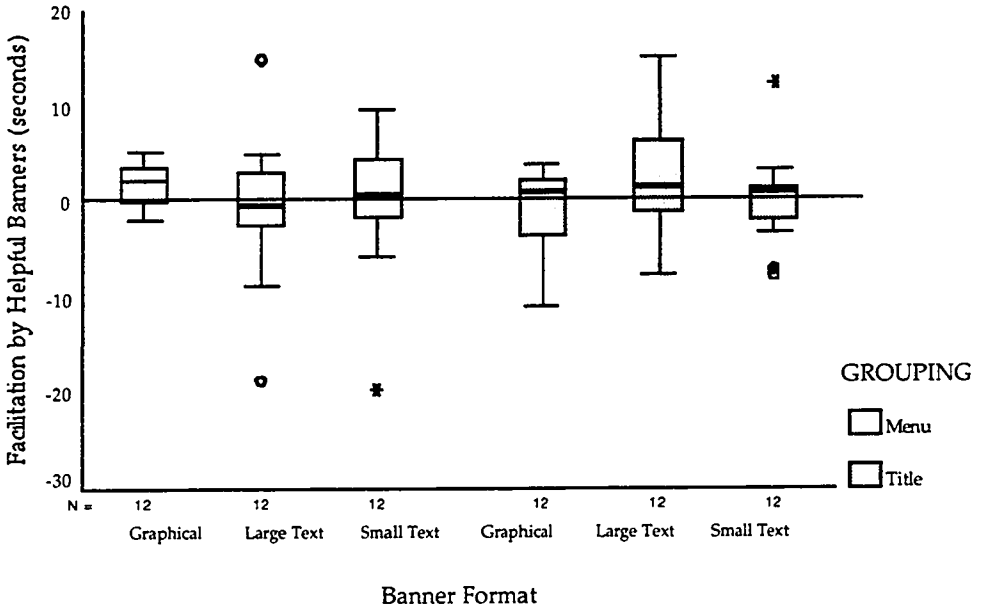


Figure 18: Difference scores in seconds for the between-subject display variables on the non-advertising trials. Positive numbers indicate faster times for the relevant banner trial.

An experimental trial took an average of 5.51 seconds to complete. The grand mean facilitation of these trials over the control trials was 0.51 seconds, which was not significantly different from zero, $t(66) = -0.19$, $p = 0.85$. This indicates that on the whole, the subjects experienced banner blindness because they were not consistently faster in selecting a category in the presence of a helpful banner. The standard deviation for the groups varied from 2.3 seconds to 8.0 seconds, but a test for heterogeneity of variance indicated this variation was not significant using the O'Brien test (Maxwell & Delaney, 1990), $F(5,66) = 1.02$, $p = 0.41$. The mean facilitation for the group with menu grouping and the graphical (ad-like) banner was 1.89 seconds. This difference was significantly different from zero, $t(11) = 2.87$, $p = 0.015$. This significance level is marginal when considered as one of six possible tests, but

still below 0.05 using a one-tailed Bonferoni-adjusted test. None of the other groups showed significant facilitation.

The differences among the groups shown in Figure 18 were tested using a mixed-factor analysis of variance with the within-subjects factor of trial type (the two control and the experimental trial types) and the between subjects factors of banner format, grouping, and block order (non-advertising trial block before or after the advertising block). With this model, a significant main effect for trial type would indicate a difference among the control and experimental trials (possibly facilitation caused by the helpful banner), and an interaction between trial type and any of the between-subjects variables would be affected by the degree of facilitation for the experimental trial varied across groups. However, there were no such significant differences. The ANOVA table is shown in Appendix B.

In the post-experiment questionnaire, subjects were asked whether they recalled seeing the relevant banner and whether they recalled seeing the irrelevant control banner. The wording of these questions is shown in Appendix A (questions 14 and 15). Fourteen subjects reported seeing the relevant banner and nine reported seeing the irrelevant banner. These positive responses came from 17 of the 71 subjects who responded to the questionnaire. Six of the respondents reported seeing both of the banners. Table 2 shows the breakdown by condition.

Table 2: Number of subjects who reported seeing a relevant or irrelevant banner. All cells had 12 participants. The title-grouping, small-text group had one subject who did not answer this question.

	Banner Format			
Grouping	Graphical	Large Text	Small Text	Total
Menu	3/12	4/12	2/12	9/36
Title	0/12	4/12	4/11	8/35
Total	3/24	8/24	6/23	17/71

All of the 14 subjects who recalled seeing the relevant (helpful) banner also reported that that banner helped them locate the correct category. Figure 19 shows the distribution of difference scores for these subjects. The mean facilitation for this group was 1.55 seconds (SD = 7.53). This group contains a large negative outlier (2.7 SDs below the mean). With this outlier eliminated, the group mean of 3.11 seconds was significantly different from zero, one-tailed $t(12) = 2.25$, $p = 0.02$. Eliminating the positive outlier shown in Figure 19 (1.8 SDs above the mean) makes the mean of 2.11 seconds significantly different from zero using a one-tailed test, $t(11) = 2.04$, $p = 0.033$.

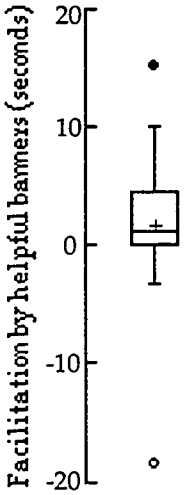


Figure 19: Boxplot of the difference scores for the 14 subjects who reported that they saw the relevant banner.

The group means and sizes for the 13 subjects (after elimination of the furthest outlier) in this "relevant-banner-seer" subset are shown in Figure 20.

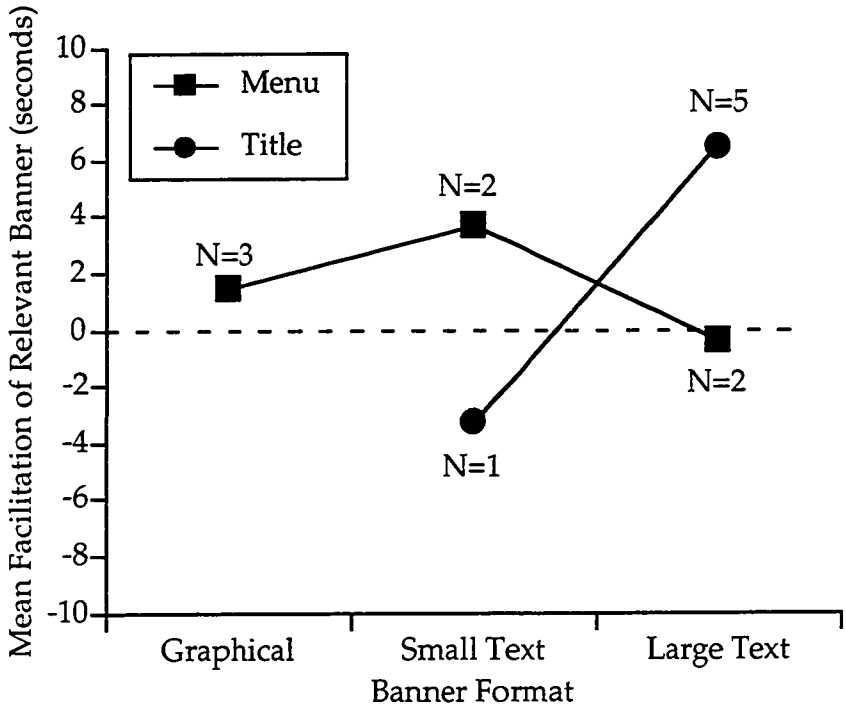


Figure 20: Group means and cell sizes for the 13 subjects who reported having seen and been helped by the relevant banner. Zero subjects in the title-grouping, graphical-format banner group reported seeing this banner.

A 2 x 3 ANOVA with SAS Type III sums of squares showed that the main effect of grouping was not significant, $F(1,8) < 0.01$, $p = 0.99$ and neither was the main effect of format, $F(2,8) = 0.36$, $p = 0.71$. The interaction was not significant, $F(1,8) = 4.21$, $p = 0.07$.

3.2.2. Advertising Trials

Banner blindness also occurred with the advertisement banners. In the recognition test, subjects were shown each advertisement that had previously been presented and were asked to indicate whether they had seen it. They responded on a four-point confidence scale: "was present," "perhaps was present," "perhaps was not present," "was not present." The subjects were also shown an equal number of control advertisements.

The experimental banners were presented in each of the factorial combinations of the four within-subjects variables: test context, grouping, animation, and block. Because grouping, block, and sometimes animation did not apply to the control advertisements, the controls were not fully factorialized. In order to match these control banner for comparison with the experimental banners, each control banners was randomly assigned for each subject to a level of animation, grouping, and block to create a full factorial design.

Across conditions, the experimental banners received confidence ratings slightly higher than the controls did. Each experimental item was matched to its corresponding control item, and a difference score was computed. The mean difference score across all conditions was 0.07 on the four-point scale, which was significantly greater than zero, $F(1,71) = 7.97$, $p = 0.006$. For all repeated-measures analyses of variance reported here, a multivariate approach was used. All repeated measures use the F test of the Wilkes' Lambda test statistic (Maxwell & Delaney, 1990).

A 2x2x2x2x3 (Experimental/Control, High/Low context test, Grouping, Animation, Block) within-subjects ANOVA was performed. There was no significant effect of block, $F(2,69) = 0.04$ $p = 0.96$ nor interaction between block and experimental/control, $F(2,69) = 2.76$, $p = 0.07$. Because one subject had a piece of missing data from one test item, analyzing the data separately for each block eliminated that line of data. Since there was no significant block effect, the data was averaged across blocks for each subject. As in the non-advertising data, the effects of interest were interactions between the experiment/control variable and the other variables. *A priori*, the confidence ratings for the control items should not vary at all based on the levels of the other variables, and thus any effect on the experimental items would produce

an interaction between the experimental/control variable and the variable that lead to the effect. However, even if the confidence ratings did vary across a variable for the control items, an interaction would still distinguish an effect on the experimental items from a response bias. There were no significant interactions of any order between the experimental/control variable and any other variables. The a summary of the results is shown in Appendix C.

There was a significant main effect for amount of context in the recognition test. High-context test items showed the actual ad banner and low context gave only the advertiser's name. The subjects were more likely to say that they had seen an advertisement when it was presented in low context (M = 2.42) than in high context (M = 2.26), $F(1,71) = 16.13$, $p < 0.0001$. But, this proved to be a response bias as the context variable did not significantly interact with the experimental/control variable, indicating that subjects were just as liberal with the low-context items when evaluating the control items as they with the experimental items.

With each low-context item, subjects were asked an additional question about which particular product was advertised. The subjects made a forced choice between two products. The mean number of correct selections made on the experimental items per subject was 0.503. This was not significantly different from chance, 0.500, $t(71) = 0.19$, $p = 0.85$. Not surprisingly, there were no significant effects in a 2 (grouping) x 2 (animation) x 3 (block) within-subjects multivariate ANOVA on the number of correct selections made for the experimental items.

In the post-experiment questionnaire, the subjects were asked whether they noticed advertisements that were shown to them during the search trials. Twenty percent said yes. Apparently, for the other 80%, the recognition

test was confusing in that they had no memory of any advertisements at all. Several subjects remarked after the experiment that they had thought that the recognition test existed in order to see whether they would begin responding that they had seen ads that did not exist under the pressure of repeated questioning. Reanalysis of the recognition test results using only the 20% of subjects who claimed to have been aware of the advertisements did not show any increase in performance. In fact, the mean difference between the experimental and control items actually decreased. The subjects' confidence in their recognition of items that were shown over controls was 0.057 points on the four-point scale. This was not significantly different from zero, $t(13) = 0.67$, $p = 0.51$. Confidence for items presented with menu grouping was greater than with title grouping by 0.15 points, which was in the predicted direction but not statistically significant, $t(13) = 0.82$, $p = 0.43$. Confidence for items presented with animation was greater than those without by 0.197 points. This difference was in the direction predicted by web design guidelines, but was not significant, $t(13) = 1.83$, $p = 0.0897$.

3.3. Discussion

The main purpose of this experiment was to investigate whether changes in how banners were presented could mitigate banner blindness. For the most part, banner blindness occurred regardless of variations on banner format, perceptual grouping, or animation. In the non-advertising case, however, there was some mitigation under the conditions where the graphical-type banner was perceptually grouped with the menu. Menu grouping was intended to mitigate banner blindness by increasing the perceptual link between the banner and the menu. That there was no overall significant main effect for grouping but a small effect for the menu/graphical

condition indicates that the effort at perceptual grouping was marginally successful.

One possible problem with Experiment 1 was that since all the banners were graphical, the banner blindness that occurred may have been caused by the subjects' avoidance of anything that resembled advertisements rather than anything perceptually distinct. The fact that the greatest facilitation in the non-advertising trials occurred in one of the two conditions that *did* use graphical, advertising-like banners indicates that banner blindness is not simply a result of the avoidance of advertisements.

Among the 14 subjects who reported seeing the potentially helpful relevant banner, there was statistically significant facilitation. This indicates that the behavioral measure used for detecting whether the relevant banner was noticed immediately (grabbed attention) was probably valid. The overall lack of strong effects in the non-ad trials, therefore, indicates that the majority of the subjects experienced banner blindness.

In the advertising trials, there did not seem to be any mitigation of banner blindness caused by changes in perceptual grouping or through the addition of small amounts of animation. The fact that there were no significant differences among conditions even for subjects who said they were aware of some advertisements may indicate that the recognition test was not sensitive enough to detect the few cases where banner blindness was less profound. The non-significant differences that did appear for this small group of subjects were in the direction predicted, though, which may indicate that under certain conditions adding animation and menu grouping may mitigate the effect of banner blindness somewhat. However, the very low rate of conscious awareness that any advertisements were present at all indicates that

most subjects experienced banner blindness across all the conditions in which the advertising banners were presented.

Experiments 1 and 2 both show that banner blindness can occur, and in some cases is quite dramatic. I now turn my attention to how to call users' attentions to important information they may be searching for without causing the ironic effect of banner blindness. Guidelines that recommend placing important items at the top of the page may be misguided. Spool, Scanlon, Schroeder, Snyder, and DeAngelo (1997) found in usability tests that users turn to navigation bars after determining that the page does not contain the information they need. At this point the user tends to have scrolled to the top or the bottom of the page. Users may start viewing the page in the center, and examine the very top and bottom only after they determine what they want is not located in the center. Therefore, it may not be better to place the "important" items at the top because the user may look there last.

A key problem with banners like those used in Experiment 2 may be the physical separation of the banners from the menu. Despite the efforts to increase the relationship between the banner and the menu by increasing perceptual grouping through background colors or greater similarity, a physical separation remained. The solution may be to indicate clearly that the "important" item is part of menu, and then to add distinguishing characteristics to make that item seem more important and draw attention.

3.3.1. Perceptual Distinctiveness

Banner blindness occurs when a web designer attempts to make a particular link distinctive on a page. In some cases, this attempt is counterproductive because although the item is distinctive, it is not noticed by some users. Basic and applied research on distinctiveness is mixed in its ability to predict this phenomenon. In this section I will describe basic

research that shows that distinctive items should be more memorable and considerably easier to spot when they are targets in visual search. However, as I will show, distinctive items in most visual search studies are distinctive on the very dimension that is relevant to the search. Therefore, much of the research in this area has limited applicability to banner blindness. *A priori*, I would assume that if a subject is shown a display of words all of which are printed in black except one, the distinctiveness of that one word should have an effect. Specifically, if the subject is asked to find the word "artichoke" and that word is the one that happens to be distinctively colored, search should be facilitated. Conversely, if the word "lettuce" is in the distinctive color, search for the word "artichoke" should be somewhat inhibited as compared to an all black display. This is not what occurs in banner blindness. The banner, being distinctive, should attract attention. For some reason, it does not. And, studies which test my *a priori* assumption are rare and have shown mixed results.

Research on the von Restorff effect shows that distinctive items are remembered better than other items that they are presented with and that perceptual distinctiveness is one way to produce this effect (Schmidt, 1991 for a review). In studies of perceptual distinctiveness and the von Restorff effect, the perceptual feature that makes an item distinctive is irrelevant to the stimulus. For example, the subject may be asked to memorize a list of words all of which are names of food. All of the items may be printed in black except one, which happens to be printed in red. The red item is more likely to be remembered than the black items, particularly in recall tests (Van Dam, Peeck, Brinkerink & Gorter, 1974). Perceptual distinctiveness has also been produced using a number of other dimensions to produce the distinctiveness including size (Karis, Fabiani & Donchin, 1984) and type style (McLaughlin, 1968).

Irrelevant perceptual features have a positive effect on a subject's ability to remember a distinctive item, so it seems reasonable to assume that distinctive items on web pages would be noticed more than other items displayed.

Studies on visual search also seem to support the idea that important items should be made distinctive. Wolfe's (1994) model of visual search, like many others, describes the existence of feature maps, which are preattentive representations of the visual field that describe the distributions of features such as color, orientation and size across the display. Wolfe's model uses the feature maps to describe patterns of activation in the display. Activation is both top-down and bottom-up. Top-down activation is guided by the searcher's goals and is used if the target is not unusual on any single dimension. Bottom-up activation is stimulus-driven. The more unique a particular object is in its context, the more likely it is that that item will attract attention to its location. A unique item will attract attention regardless of what the viewer is actually looking for (Wolfe, 1994). Therefore, a red item surrounded by blue items will attract a greater amount of activation than the blue items regardless of what the viewer is looking for.

Web users usually search for semantic information. They may search for tips on dog grooming, for example. If a web designer provides tips on dog grooming, the existence of bottom-up processing suggests that if the link to the dog grooming information happens to be in large type, or on a brightly colored background, the person seeking dog-grooming information might find it very quickly. And, if another person is looking for information on pet food, the existence of the large, brightly colored link to dog grooming might attract the attention of the pet-food searcher as well, and perhaps cause the search for pet food to take a little longer. The fact that the dog-grooming link

is large and brightly colored is irrelevant to the search, but might affect it anyway.

Most empirical work on visual search does not address the effect of irrelevant perceptual distinctiveness on search speed. Treisman has done a series of studies wherein the search target either differs from distracters on a single dimension or on a conjunction of two dimensions (e.g. Treisman & Gelade, 1980; Treisman & Sato, 1990). Her feature integration theory holds that when the target has a unique feature not shared by distracters, search can be parallel and quick. When the target is only unique on a conjunction of features, search is serial and slower (Treisman, 1982; Treisman, 1991). It may seem that feature integration theory predicts that distinctive items, those which require only a single dimension to be considered perceptually unique, would always be found more quickly than other items. However, feature integration theory is based on research in which the target is always defined by the same perceptual dimension which make it unique. Therefore, feature integration theory says nothing about cases in which a searcher is searching for a particular word, letter, or semantically defined target which is incidentally distinctive on some other dimension.

A competing theory may seem to have more relevance to incidental distinctiveness. Duncan and Humphreys (1989) proposed that difficulty in search is purely a function of the amount of similarity between the target and the distracters, and the similarity of distracters to one another. The more similar a target is to its distracters, the harder it is to find, and the more the distracters resemble one another, the easier the target is to find. Therefore, it would seem that if there is one perceptually distinctive item in a field of similar items and the target is the distinctive item, search for that item should always be faster. And, if the target were one of the similar items,

search would be slower both because of its similarity to many of the other items and because of the dissimilarity of the one distinctive nontarget to the other nontargets. However, Duncan and Humphreys' theory is based on the idea that a searcher compares a template of the searched-for item to the items available on the display. The template contains only the information which is needed to distinguish targets from nontargets. Therefore, this theory also provides little predictive power when the perceptual distinctiveness is irrelevant to the search. If a person is searching a list of words printed in black for the word "apple," the template would contain only the information for the spelling of the word apple and no information about, for instance, the color of the word. Therefore, if the target word happens to be printed in orange, Duncan and Humphreys' theory on the surface would say that the color decreases the similarity between the target and nontargets. But, applying the theory more completely indicates that the difference may not aid in the search process because color information would not exist in the template for the word "apple."

Researchers in the applied domain have varied the existence of features to increase the perceptual distinctiveness of targets which are defined without reference to dimensions like color or size. For example, Nygren (1996) had subjects search lists of words for a particular target word. The target was made distinctive using color, shade, spacing, italics and size. The targets were found much more quickly than if none of the items in the display were made distinctive. However, distracter items were never distinctive and so the subjects were able to learn that distinctiveness should always be attended to. There were no differences among the five dimensions used to create the distinctiveness.

Fisher and Tan (1989) performed a similar manipulation where subjects were asked to search for a one number among four number distracters. In some cases all items were perceptually similar, and in other cases the target was made distinctive using color, reverse video or blink. Distinctive targets were found most quickly when they were a different color than the distracters, and there was equal facilitation for reverse video, but only when the target was located close the point of fixation. Making the target distinctive by making it blink did not facilitate search. This experiment and the Nygren study show that when the target is distinctive it is found more quickly for most dimensions. However, in these two experiments, distracters were never distinctive. Therefore, subjects could actually search for the "odd man out" rather than the particular word or number they are asked to search for, and this would facilitate search (Ullman, 1984).

Other studies have examined "coding" of portions of the display by changing perceptual elements. For example, Smith and Goodwin (1971) coded portions of displays by making them blink. The targets were specific instances of classes of objects, and there were five classes. In most trials a whole class of objects would blink and the subjects were told what type of blink coding would be used. Since the subjects knew whether or not the target was blinking, the blink status was relevant to the search and in fact the searchers were performing a conjunction search. In addition, since one fifth of the items on the screen blinked at a time, blinking did not make particular items distinctive. In one condition, half of the items blinked and the blink status of the target was random. In this case, blinking was irrelevant to the target, but blink still did not provide distinctiveness. The results were that when classes of objects blinked or all objects blinked except the target class and the subject knew the trial type, the coding provided faster search than when nothing

blinked or everything blinked. When the blink coding was random, however, there was no facilitation. This showed that subjects could effectively use blink as a cue to focus their attention on subsets of objects, but it does not shed light on the effects of distinctiveness.

A study by Fisher, Coury, Tengs, and Duffy (1989) was similar to the blink coding study, but the probability that the target would be included in the coded set was varied, as well as the distinctiveness of the coding. The subjects were asked to search for a target word among 35 other words. The coding was done with color. The "validity" of the coloring was varied, as well as the number of colored items. In the low validity case, there was a 25% probability that the target would be in the unusual color, and in the high validity case, the probability was 75%. The number of items that were displayed in the unusual color was either one, three, six, or twelve. The subjects were informed before each trial how many items would be colored, and of the validity level. Therefore, the color dimension was relevant to the search. The results were that under both levels of validity, when the target was colored, search was faster than when no words were colored. The level of facilitation was higher for the higher level of validity. However, in cases where six or twelve of the items were colored, coloring was not particularly distinctive. In the cases where one or three items were colored, search was faster when the target was colored. If the distinctiveness of those items was involuntarily attracting attention, then there should be interference when the target was not colored and one or three distracters were. However, in these trials the search time was very close to and in some cases somewhat faster then in plain trials. Distinctive features did not seem to pull the searchers attention to them when the subjects were informed of the trial type.

van Nes, Juola and Moonen (1987) performed a similar manipulation but they did not inform the subjects of the target color probabilities. In this study the displays were paragraphs of words and subjects were asked to search for particular words. The displays were in one, two, or four colors, and distinctiveness was varied although coloring validity was not. One group of subjects was told both the target word and its color before each trial. Therefore, these subjects were performing conjunction searches and color validity was 100%. For these subjects, the more words that shared the target's color, the slower the search. For the subjects who were not told the target color, search for targets was faster than the single-color display only when the targets were in a rare color. They were much faster than the single-color time when only two colors were used and the target was in a color used on only 10% of the words. They were marginally faster than the control when two colors were used and 25% of the words shared the target color and when four colors were used and 10% of the words shared the target color. In all other conditions, search was slower than the control. Search was particularly slow when two colors were used and the target shared a color with 90% of the other words, and when four colors were used and the target shared color with 25%, 75% or 90% of the other words. Except for the 4-color, 25% condition, these are all cases where the distracters took rare colors. Therefore, the rare colors did seem to draw the subjects' attention and facilitate performance when the target was distinctive and inhibit performance when distracters were distinctive. This occurred even though color instructions were not provided to the subjects. The level of color validity was 50%, because there were an equal number of trials where the target appeared in rare colors as in common colors. Therefore, the subjects may have implicitly learned that

searching the distinctive areas first was a good strategy because the target appeared there half the time.

The attentional attraction of rare colors was validated through a small eye-tracking study reported in the same paper. van Nes et al. found that some searchers used a "reading" strategy, scanning the paragraph for the target from top to bottom, and others used a "hopping" strategy, skipping around the paragraph. Both strategies were susceptible to attraction to rare-colored items. The readers showed several saccades that jumped out of the reading pattern and to the distinctive items. The hoppers showed even more attraction to the distinctive items, saccading to most of these before examining other parts of the display.

So van Nes et al. found that distinctive coloring attracts attention for subjects who are not explicitly instructed about the manipulation, at least when the validity of using the distinctive colors is around 50%. Gomberg (1985) found the opposite, using other dimensions to create the distinctiveness. Gomberg created displays of five numbers in which the subject had to identify which member of the target set (the digits 1, 2, 3, 4) existed among four distracters (from the set 5, 6, 7, 8, 9). In control trials, all items were in similar type styles. In other trials, Gomberg made one of the five items distinctive by placing a box around it, by making it blink, or by displaying it in reverse video. The target item was the distinctive item in 50% of these trials, so the validity level was 50%, just as it was in study by van Nes, et al. Gomberg found that the trials that contained distinctive items were as a whole slower than the control trials. Also, target-distinctive trials were very slightly, but not significantly, slower than distracter-distinctive trials. Tan and Fisher (1987) replicated Gomberg's results. This study showed that in all distinctive trials, regardless of the dimension used (boxing, reverse video,

blink) or whether the target or distracter was distinctive, the control was faster.

Fisher and Tan (1989) hypothesized that no facilitation was found because boxing, reverse video and blinking were not the appropriate perceptual dimensions. They repeated the study, maintaining the 50% validity level and creating distinctive items through color, reverse video, or blinking. As in the other experiments, distinctive trials were slower overall than the control trials. Unfortunately, Fisher and Tan do not report detailed analysis comparing target-distinctive and distracter-distinctive trials. They do report that in target-distinctive trials there was facilitation for color over the control and other dimension trials, but the magnitude of the facilitation is small and significant at only some target positions. From the graphs in the paper, it appears that the reverse-video distinctive targets were similar to control trials, and that blink was considerably slower than any other condition when the target was near the point of fixation. As for the distracter-distinctive conditions, statistical analysis is not reported but it appears that there was inhibition for all the dimensions and the control trials were the fastest. When Fisher and Tan raised the level of validity to 100%, there was facilitation for the color and reverse-video conditions, particularly when the target was near the point of fixation. Blink trials were still slower than controls.

Tan and Fisher (1987) also looked at the effect of display size. Using only color to create distinctiveness, they used displays that had one row of five numbers, as in the other work described above, and added displays that had five rows of five numbers. Validity was either 50% or 90% and these trials were blocked. The overall effect of distinctiveness was still negative for the small displays, but was now faster for the larger displays, particularly

when the validity was 90% (and thus there were many more facilitation trials than inhibition trials). When they considered only the target-distinctive trials and the middle-row of the larger display, there was equal facilitation for both display sizes and both levels of validity, although the 90% validity trials were faster than the 50% trials. Also, there were significant order effects, so that subjects who saw the 90% validity block first had facilitation in the both the 90% and 50% blocks, but those who began with the 50% block had facilitation in the 90% block only. Therefore, the attraction to the distinctive item seems to be more of a learned advantage than a preattentive attraction to the "odd man out."

In summary, it is unclear whether perceptual distinctiveness attracts attention when it is irrelevant or marginally relevant to the task. A number of screen design guidelines suggest that making a particularly important item different from the other choices, by putting a box around it, making it larger, or making it a different color, should help attract attention to it (Detweiler & Omanson, 1996; Fleming, 1997; Shneiderman, 1992). Empirical research has not been as clear. When the distinctiveness is highly relevant or completely redundant, it aids search, as has been found in basic visual search studies (Duncan & Humphreys, 1989), in searches for words in large sets of distracter words (Nygren, 1996; van Nes et al., 1987), and in small sets of single digits (Fisher & Tan, 1989; Tan & Fisher, 1987). In cases where the distinctive feature is relevant only 50% of the time, distinctiveness has been shown to facilitate target-distinctive searches and inhibit distracter-distinctive searches for words in long paragraphs (van Nes et al., 1987). But, in searches for single digits in small or medium-size displays, the results are much less clear-cut and facilitation for target-distinctive searches is rare (Fisher & Tan, 1989; Tan & Fisher, 1987).

Experiments 1 and 2 showed that one way of making targets distinctive, by separating them from the menu, inhibits search for those items and causes banner blindness. The next two experiments place distinctive targets within the menu and emphasize them using color or text size. These studies attempt to answer two main questions. First, will adding emphasis to an item within a menu cause banner blindness? And second, if this emphasis does not cause banner blindness, will the emphasis cause the users to select the emphasized item?

4. EXPERIMENT 3

4.1. Method

4.1.1. Participants

Thirty three Rice undergraduate psychology students participated in this study in exchange for course credit. The mean age was just under 20 years, and 19 of the subjects were female. Thirty of the subjects used the web every day or several times a week. The others used it at least once a month.

4.1.2. Procedure

The subjects were given eight tasks to complete. The tasks were printed on index cards and required the subject to look for fairly specific information. The tasks are shown in Appendix D. For each task, the subjects were asked to type search terms into a field on the screen shown in Figure 21. The subjects were told they could enter the terms in any format they wished: one term or several, spaces as separators or commas, etc.

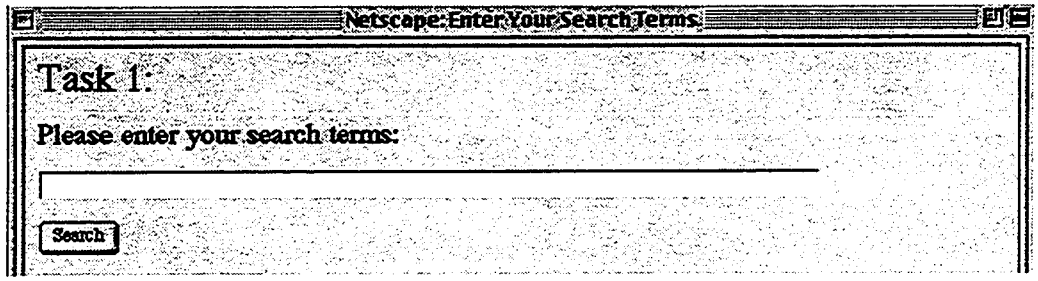


Figure 21: Subjects entered search terms for each task.

After clicking the search button, the subjects were shown a screen with eight "hits" related to the search task. The "hits" provided were the same for all subjects, regardless of the search terms entered. A "hit" page is shown in Figure 22.

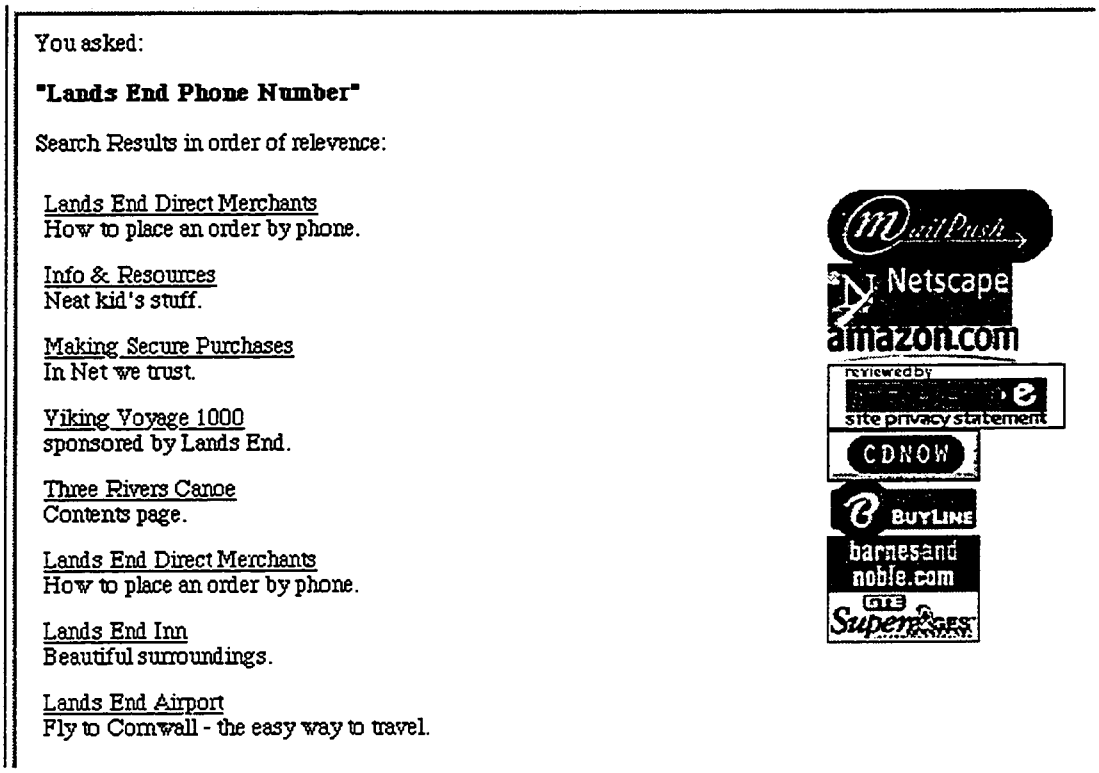


Figure 22: A "hit" page for one of the control tasks. The underlined text items were blue links. The remaining text was black.

The subjects were instructed to select whichever link on the hit page they thought would best lead them to find the information described in the task. They were also instructed to take their time in the selection, and not to

rush through the task. After they made the selection, the task was over, the subjects were taken immediately to the search page to begin the next task.

After the eighth task, the subjects filled out an on-line survey asking them how frequently they use the web, their gender, and their age. They were also given the opportunity to type comments to the experimenter.

4.1.3. Materials

In each non-buffer trial, six of the links provided on the hits pages were peripherally related to the task, but were fairly clearly "wrong" in that they probably would not yield the information the subjects were asked to find. A seventh link was fairly clearly the "correct" link. On all the experimental and control trials, this link was displayed twice on the hit page. It appeared in the first position in the list and also in either the fifth, sixth, or seventh position. Figure 22 shows a control trial where the "correct" link appears in positions one and six.

In the experimental trials, one of the two correct links was emphasized to make it visually different from the other links. Two types of emphasis were used, color and size. Color-emphasis items had a bright yellow background. Size-emphasis items were printed in large text. Subjects completed three experimental trials. There were two color-emphasis trials. In one, the top-color condition (Figure 23), the first hit in the list was emphasized with color. In the middle-color condition (Figure 24), the lower correct hit (in position five, six, or seven) was emphasized with color. The size trial used size emphasis in the first hit, as shown in Figure 25.

You asked:

"CEO of Coca Cola"

Search Results in order of relevence:

Top Executives at Coca Cola
Profiles of the CEO, VPs.

The CocaCola job seekers page.
How to ace an interview.

Coca Cola not really a secret.
Email me for the formula.

Chris's Coca Cola Page
Americans drink 7000 Cokes a day!

Endorsements for JAMMINGS
#1 authority on business creativity.

India Current Events
Coke uncorks plan for bottlers.

Top Executives at Coca Cola
Profiles of the CEO, VPs.

How to choose your successor
Positive leadership newsletter.

Figure 23: A top-color experimental-trial hit page. The emphasized link had a bright yellow background where gray is shown above.

You asked:

"CEO of Coca Cola"

Search Results in order of relevence:

Top Executives at Coca Cola
Profiles of the CEO, VPs.

The CocaCola job seekers page.
How to ace an interview.

Coca Cola not really a secret.
Email me for the formula.

Chris's Coca Cola Page
Americans drink 7000 Cokes a day!

Endorsements for JAMMINGS
#1 authority on business creativity.

India Current Events
Coke uncorks plan for bottlers.

Top Executives at Coca Cola
Profiles of the CEO, VPs.

How to choose your successor
Positive leadership newsletter.

Figure 24 A middle-color experimental-trial hit page.

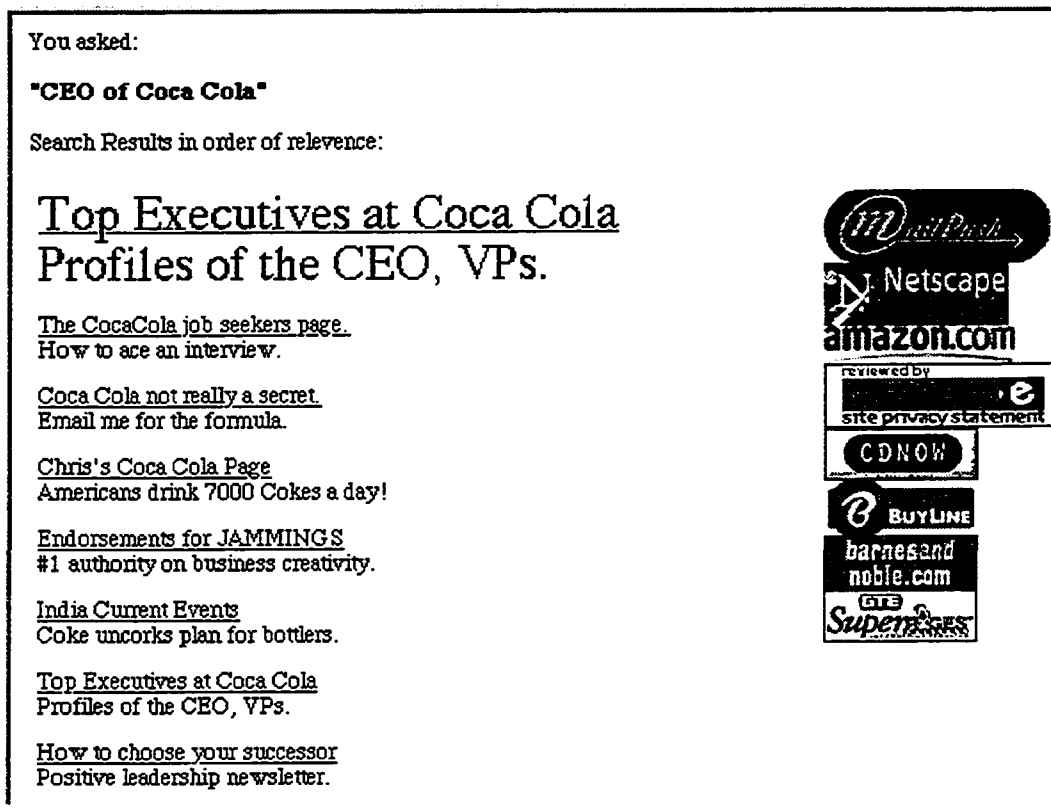


Figure 25: A size-emphasis experimental-trial hit page.

The experimental trials occurred in trials four, six, and eight. Control trials, with no emphases, occurred in trials three, five, and seven. Trials one and two contained only one correct hit, and it occurred in position two or three. Table 3 shows a summary of this design. The order of the three types of experimental trials was counterbalanced across subject in a Latin Squares design.

The subjects were not given any feedback as to whether they had clicked a correct link or not.

Table 3: Trial types and locations of the links subjects were expected to click within the eight-item hit list.

Trial	Type	"Correct" link locations
1	Buffer	2
2	Buffer	3
3	Control	1, 6
4	Experimental	1, 7
5	Control	1, 5
6	Experimental	1, 6
7	Control	1, 7
8	Experimental	1, 5

4.2. Results

Subjects made no errors in hit selection on the experimental and control trials. Two subjects each made one error on the buffer trials.

Each non-buffer trial was dummy coded with a "1" if the subject clicked the top link and a "0" if not. The scores for the three control trials were averaged together and ranged from zero to one. Figure 26 shows the mean control score computed in this way and the percentage of subjects who clicked the top link (scored a "1") in each of the experimental conditions. Differences between control scores and experimental scores were tested for significance using two-tailed Dunnett's t-tests.

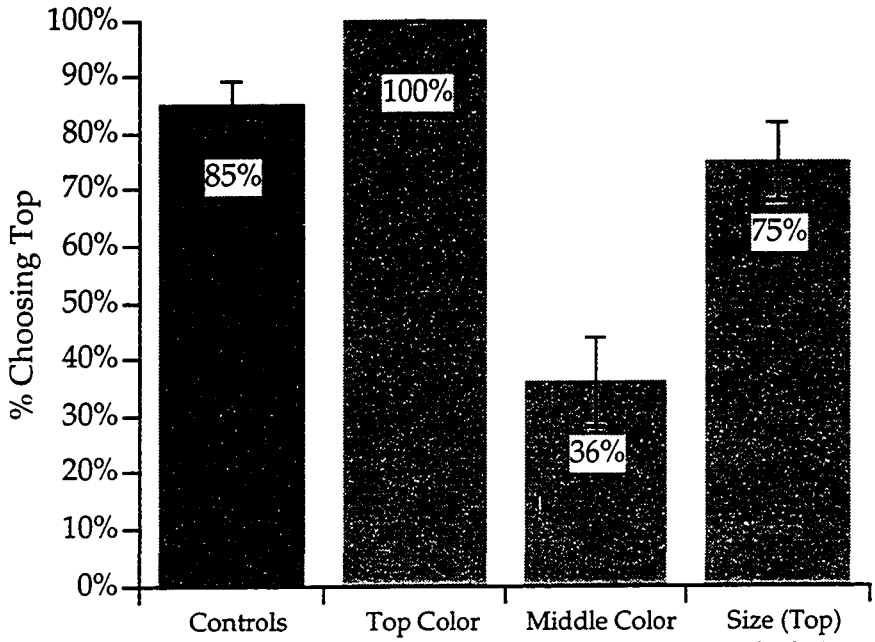


Figure 26: Percentage of subjects who selected the top link in each type of trial. For the controls, the number of times the top link was chosen across the three control trials was averaged for each subject. Error bars indicate standard error of the mean.

In Figure 26, the dependent variable is the percentage of subjects who selected the top link. In the size and top-color conditions, higher scores indicate that more subjects selected the emphasized link. In the middle color condition, a lower score indicates that more subjects clicked the emphasized link because the emphasized link was in the middle, not at the top.

When the top choice was emphasized with color, it was selected by 100% of subjects. This rate was significantly greater than selection of the top link on control trials, Dunnett's $t(32) = 3.24$, $p < 0.01$. Similarly, when an item in the middle of the list was emphasized using color, the identical top item was selected less often than it was in the control trials. This difference was significant, $t(32) = -5.39$, $p < 0.01$. When the top item was emphasized with text size, it was selected slightly less often than the top item was in control trials but this difference was not significant, $t(32) = 1.14$, $p > 0.05$.

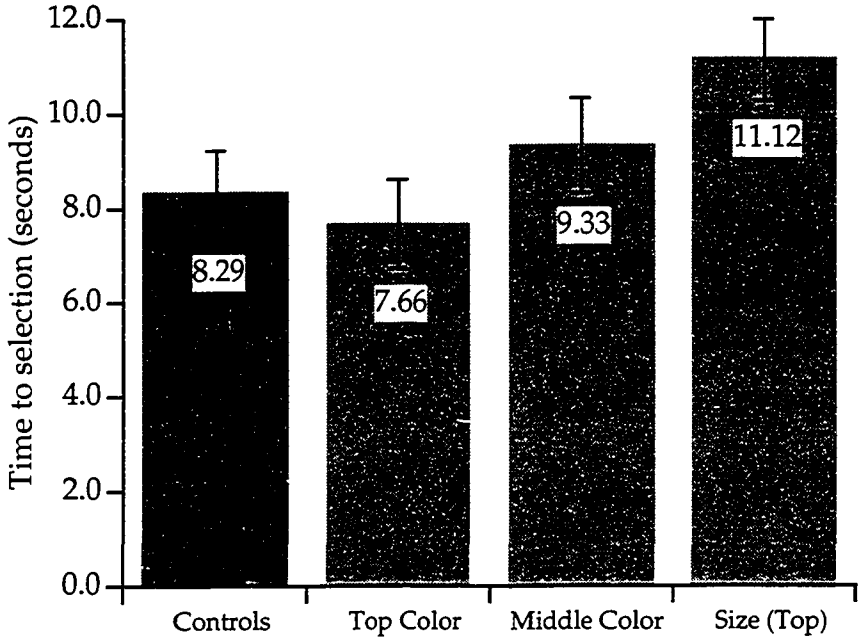


Figure 27: Mean time to selection for each time of trial.

The time it took the subjects to make their selection showed a different pattern (See Figure 27). The three selection times for the control trials were averaged together for each subject and compared to the experimental trials. Paired Dunnett's t-tests were used again to test for significance. The time it took to make a selection was slightly shorter in the top-color emphasis condition than in the controls but this difference was not significant, $t(32) = -0.73$, $p > 0.05$. The time it took to make a selection in the middle-color condition was slightly longer than in the control condition but this difference was not significant, $t(32) = 0.94$, $p > 0.05$. However, the time to selection for the size-emphasis condition was considerably longer than the controls and was significantly different, $t(32) = 2.90$, $p < 0.05$.

4.3. Discussion

Color highlighting does appear to influence users to select the emphasized link. There does not seem to be any banner blindness associated with adding this type of salience to a piece of text in a list. Users were not only were unlikely to miss the color-emphasized items, they tended to choose the

colored item over the other items. The existence of color emphasis did not seem to either slow down or speed up selection by a appreciable amount as compared with controls.

The color-emphasis items, unlike the banners in Experiment 2, looked like they were a part of the menu on the "hit" page. I originally expected that the size-emphasis used in the current experiment might cause banner blindness because it might look more like a title than part of the menu. In fact, most subjects noticed it as evidenced by the fact that 75% of them selected it. Although this selection rate was lower than the selection rate for the item in the same position in the control trials, it was not significantly different. However, the size emphasis did seem to cause a negative effect in that subjects were slower to make a selection when it was present. Perhaps subjects didn't notice the item at first glance and only recognized as part of the menu after a few seconds. If so, this could be considered a form of banner blindness in that the very large text did not cause the item to "jump out" at the users and in fact may have caused some inhibition of that item, although the effect was temporary.

The selection from the list of hits in this experiment was somewhat artificial because the search always returned two identical items that were very obviously the best choices for the task. In fact, a number of subjects commented on this. In effect, the subjects were only required to make a simple choice between two identical items. It is still unclear whether color emphasis would encourage users to select a emphasized item over another that was equally good, but different.

Another unresolved question is whether the first item in the list is in a privileged position. The design of the current experiment and the fact that subjects were encouraged to select items earlier in the list (because the list was

introduced by "Search results in order of relevance:") makes it difficult to tell whether web users favor items at the top. It could be that emphasis has a greater effect lower in the list because those items are otherwise undistinguished, and thus emphasis has an incrementally larger effect.

5. EXPERIMENT 4

In this experiment I examined the effect of emphasis when there were multiple "good" choice available that could provide the answer to the task. In addition, I manipulated the location of the emphasis (top versus middle of the list) more thoroughly than in Experiment 3. Color emphasis was used again, and was compared to size emphasis. Because size emphasis using very large text seemed to cause some negative effects, I looked at text that was only marginally larger than the other text. Perhaps because this slightly-larger text was barely noticeable, it would influence users to favor it even if they didn't explicitly notice what made it different.

5.1. Method

5.1.1. Participants

Ninety-three Rice undergraduate psychology students participated in this study in exchange for course credit. The mean age was just over 19 years, and 56 of the subjects were female. Eighty-four of the subjects used the web every day or several times a week. The others used it at least once a month. One subject did not complete the survey, and thus gender, age, and web use for that subject is unknown.

5.1.2. Procedure

The procedure was essentially the same as in Experiment 3. The subjects were given search scenarios on index cards. They typed in search terms, and then selected the best link from a list of eight on the "hits" page.

After completing nine tasks, they filled out the same on-line survey as in Experiment 3.

5.1.3. Materials

The text of the nine search tasks is shown in Appendix D. The first eight tasks were the same as in Experiment 3. The format of the hit pages was similar to Experiment 3, but varied in a few ways. The notation that the search results were listed "in order of relevance" was removed. Also, the text content of the hits changed. In the experimental and control trials, there were five obviously "wrong" links that were related to the search task on a surface level, but would clearly not yield the information the subjects were looking for. The other three links were all "correct" in that they could reasonably be assumed to lead to the information asked for in the task scenario. I pre-tested these three "correct" links by having graduate students rank order them by what they would select if completing the tasks. The resulting choices used in the study were not assumed to be exactly equal, but all had been selected as a first choice by at least some of the pre-testing graduate students.

Table 4 shows the order of the trial types and the locations of the correct links in the hits lists. For the emphasis and distracter trials, the locations of the "correct" link texts were counterbalanced across subjects so that each specific link content appeared in each position (top, middle, or bottom) an equal number of times. Trials one, two, and six were buffer trials. In these trials, position one was never used for correct links, and position two was always used for correct links, so as to mitigate the effects of location priming of position one and inhibition in position two that might occur from the pattern used in the other trials.

Table 4: Trial types and locations of the links subjects were expected to click within the eight-item hit list. For subjects in the control group, emphasis and distracter trials looked the same.

Trial	Type	"Correct" link locations
1	Buffer/Distracter	2, 3, 4, 6, 8
2	Buffer/Distracter	2, 4, 5, 6, 7
3	**Emphasis**	1, 3, 5
4	Distracter	1, 5, 7
5	**Emphasis**	1, 4, 8
6	Buffer	2, 3, 5, 7, 8
7	**Emphasis**	1, 5, 8
8	Distracter	1, 3, 7
9	**Emphasis**	1, 4, 6

Thirty-three subjects were assigned to the control group and 60 to the experimental group. For the control group, all the links on all the pages were in plain text, with no emphases. For the experimental group, there was emphasis on one hit in trials three, five, seven, and nine. There were two emphasis variables: location and format. Emphasis was placed on either the first item in the list or the middle "correct" link, which was located in position three, four, or five (see Table 4). The two formats were color, which was created with a bright yellow background as in Experiment 3, and size, which was different from Experiment 3. In the current experiment, size emphasis was created by using a font only slightly larger than was used for the other text on the page. Size emphasis is shown in Figure 28. The four emphasis conditions (position x format) were counterbalanced for order in a Latin Squares design across the four emphasis trials.

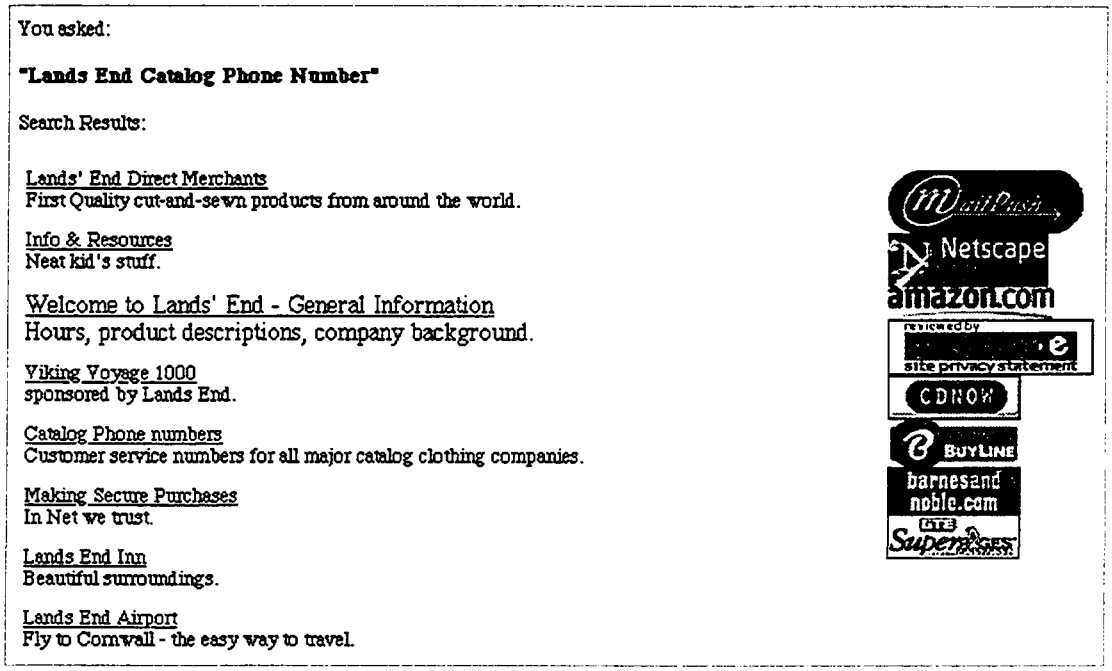


Figure 28: A hit page with barely-noticeable size emphasis on the middle position, third in the list.

5.2. Results

One subject made an error on a non-buffer trial, but noted in her comments at the end of the experiment that she had made her selection in that trial accidentally. The subject was replaced. Otherwise, performance was error free outside of the buffer trials.

Because there were three possible link locations to choose from, and because the content of the links was counterbalanced so that each link appeared in each location an equal number of times, theoretically, each position (top, middle, bottom) in the hit list had an equal chance of being selected in the control condition. However, it is possible that even with no emphasis subjects would be likely to pick a link in one position (perhaps the top) over the other positions. In fact, the data suggest that no position was privileged over the other positions. The control data from the four trials corresponding to the emphasis trials in the experimental group (trials three, five, seven, and nine) were evaluated to determine whether subjects were

more likely to choose a link in one position over the others. I computed the proportion of times each subject clicked each position across the four trials. Averaging across subjects, the top position was selected 34.09% of the time, the middle position was selected 35.61% of the time and the bottom position was selected 30.30% of the time. A one-way within-subjects ANOVA showed no significant effect of position Wilks' Lambda $F(2,31) = 0.27$, $p = 0.76$. More liberal t-tests comparing the top position to the average of the other two positions, ($t(32) = 0.15$, $p = 0.88$), the middle position to the average of the other two positions ($t(32) = 0.48$), $p = 0.64$) and the bottom to the average of the other two positions ($t(32) = 0.73$, $p = 0.47$) also found no significant differences among the positions in the control group.

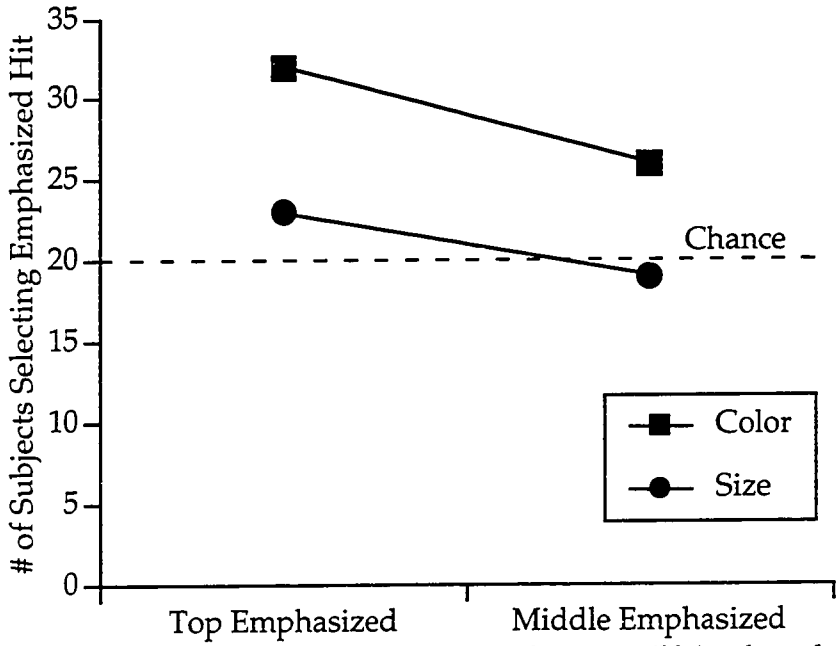


Figure 29: Number of subjects (out of 60 possible) who selected the emphasized link for each type of trial. The dotted line indicates chance level.

Because the performance of the control subjects was so close to the theoretical norms expected, the experimental data from the highlighting trials were compared to expected values of 33% for each position. Color emphasis in the top position in the list had the greatest influence on users (see Figure

29). Thirty two out of 60 experimental subjects (53.33%) chose the emphasized item in this condition. This was significantly greater than the theoretical expected value of 33%, $\chi^2(1) = 10.12$, $p = 0.0015$.

As shown in Figure 29, a plurality of subjects also selected the emphasized item when it was emphasized by color in the middle position. In this condition 26 subjects, or 43.33% selected the emphasized item, but this was not significantly greater than chance, $\chi^2(1) = 2.59$, $p = 0.11$.

Size emphasis had less of an effect. With size emphasis in the top position, the emphasized item was selected by 23 subjects (38.33%) which was not significantly different from chance, $\chi^2(1) = 0.66$, $p = .42$. With size emphasis in the middle position, the number of subjects who selected the emphasized item was nominally below chance, 19 subjects (31.67%), $\chi^2(1) = 0.08$, $p = 0.78$.

Overall, the time to make selections from the hit lists took considerably longer in this experiment than in Experiment 3. The average time to selection for the control subjects was 20.75 seconds, and for experimental subjects was 21.60 seconds. There was no significant difference between the groups, $t(91) = 0.45$, $p = 0.65$. The overall time to make a selection did not vary much among conditions, but an interesting pattern emerges when the selection times are sorted out by which of the correct items (top, middle, or bottom) were selected. Figure 30 shows the amount of time taken to make a selection as a function of the position of the item selected.

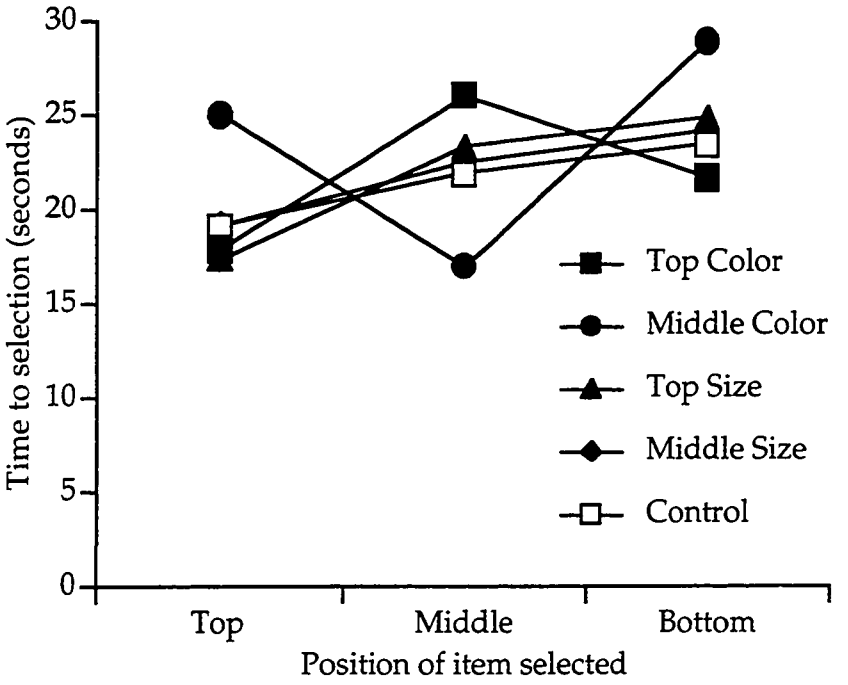


Figure 30: Time to selection under each condition as a function of the position of the item selected.

As shown in Figure 30, time to selection for the control subjects was very similar to that of experimental subjects in the size-emphasis trials, with a slight linear increase the further the selected item was from the top of the list. The two color-emphasis conditions, however, show quadratic components of trend.

In order to statistically compare the control condition to the color-emphasis conditions, one of the four control trials was selected at random from each control-subject's data. The four control trials for each subject could not simply be averaged together because not all subjects made selections from all positions. Randomly selecting one trial for each subject makes the data congruent to the emphasis data and ensures that each subject contributes exactly one data point to the analysis. After the random selection the resulting cell sizes are shown in Table 5.

Table 5: Number of subjects per cell used in analysis of time-to-selection as a function of condition/group and position of the item selected.

	Top	Middle	Bottom	Total
Top Color	32	18	10	60
Middle Color	23	26	11	60
Control	9	14	10	33

The time to selection for all conditions was positively skewed. Natural log transformations on the time data was used to create normal distributions. A group (2) x position (3) between-subjects ANOVA comparing the top color condition to controls across the three positions shows an insignificant main effect for group, $F(1,87) = 0.32$, $p = 0.57$, an insignificant main effect for position, $F(2,87) = 1.55$, $p = 0.22$, and an insignificant interaction, $F(2,87) = 0.86$, $p = 0.43$. The quadratic trend for the top-color condition was not significant, $F(1,87) = 2.12$, $p = 0.72$.

A 2 x 3 ANOVA comparing the middle color condition to controls across the three positions showed a significant main effect for position, $F(2, 87) = 3.83$, $p = 0.025$ and no significant effect for group, $F(1,87) = 2.08$, $p = 0.15$. The interaction was not significant, $F(2,87) = 2.25$, $p = 0.11$. The linear trend for the control data alone was not significant, $F(1,87) = 0.54$, $p = 0.47$, but the quadratic trend for the middle-color condition was significant, $F(1,87) = 15.23$, $p = 0.0002$. Therefore, the time to select the middle item when it was colored was faster than selecting the top or bottom item.

5.3. Discussion

In this experiment users were presented with making a judgment between three reasonably good choices. For the experimental group one of the choices was emphasized with color or size. Size emphasis had virtually no effect with behavior in trials with size emphasis being very similar to controls

in both which items were selected and in latency. Color emphasis exerted some influence on the users' choices, although to a lesser extent perhaps than it did in Experiment 3, when the presence or absence of color was the only distinguishing characteristic.

5.3.1. Interaction Between Highlighting and Position

Where should a designer place the most important item in a list? The data from the control group suggest that, at least in a short, non-scrolling list, it doesn't matter. Based on the criterion of which item was selected, there were no differences among positions in the absence of any emphasis. However, placing an item first in the list and highlighting it effectively (color was effective, text size was not) significantly increased the likelihood of that item being selected. Color highlighting elsewhere in the list did not reliably increase the selection rate of the colored item. But, color highlighting in the middle of the list did reliably have an effect on the secondary dependent variable, time to selection. Therefore, this configuration does seem to draw the attention of searchers.

Color in the middle of a menu has some intriguing possibilities. Adding highlighting to the middle of a menu seemed to change how the menu was read and used. These results are repeated in Figure 31. None of the trends besides the quadratic trend for the middle-color condition were significant, but the data suggest a possible model for further investigation.

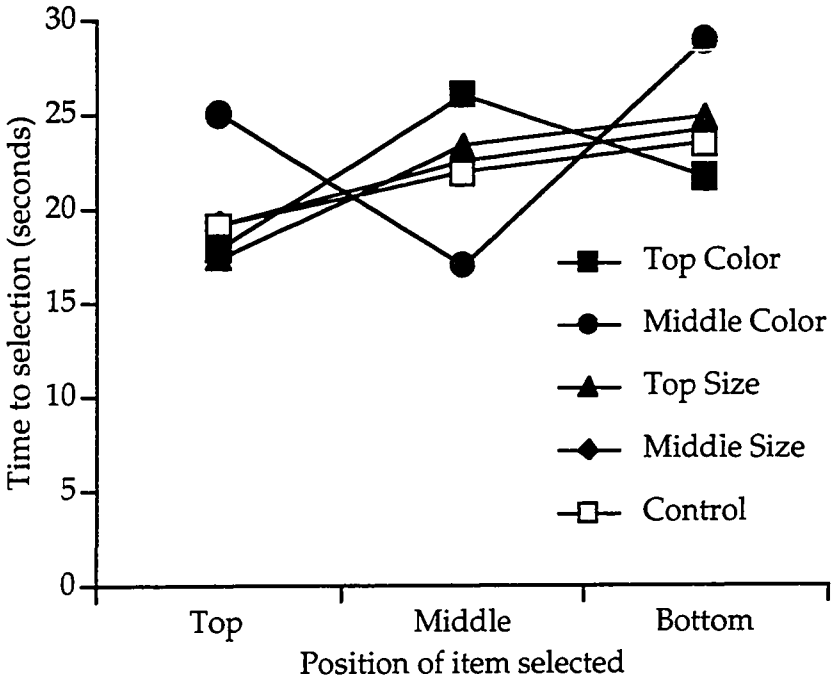


Figure: 31 Time to selection under each condition in Experiment 4 as a function of the position of the item selected. Repeat of Figure 30.

Most menu-selection studies find that the further a target item is from the top of the menu, the longer it takes to select it. This serial position effect is similar to the nonsignificant linear trends shown in Figure 31 for the control condition and the two size conditions. The serial position effect could indicate a systematic search from top to bottom (Lee & MacGregor, 1985; Norman, 1990). But, modeling of empirical data suggests that menu searchers use a combination of parallel and serial processing of items, scanning the items systematically top-to-bottom with some random hops (Hornof & Kieras, 1997). An eye-tracking study of menu selection indicates that users search menus in upward and downward "sweeps." Search almost always starts with a long downward sweep, followed by alternating upward and downward sweeps that cover more items in less time. Sometimes users use the upward sweeps to move the eye for a repeated top-to-bottom scan and sometimes they alternate between reading items top-to-bottom and bottom-to-top (Aaltonen,

Hyrskykari & Räihä, 1998). Similar results (but for much simpler menus) were found in another eye tracking study by Byrne, Anderson, Douglass, and Matessa (1999).

Because the serial position effect shown in the control data in Figure 31 looks similar to the effect found by Aaltonen et al. in the eye-tracking study, it seems reasonable to interpret the results assuming the subjects used similar up-and-down-scans of the available options. Some subjects probably read first from the top to the bottom and then scanned back up to their preferred option, while others did repeated scans from the top down. But the trials were color highlighting was used probably resulted in different scan patterns.

Subjects who selected the highlighted item when it occurred in the middle position did so quickly - in about the same time it took to select the top item under other conditions. And, subjects in the middle color trial who rejected the highlighted item took longer to make their selections. Perhaps highlighting an item in the middle disturbs the typical up-and-down sweeps and causes users to focus on the highlighted item first. Therefore, the read order is no longer from top to bottom, but from middle to top to bottom or some other order. Therefore, latency for selection of the middle item is similar to selection of the top item under other conditions, and the time to select the top or bottom item is proportionally slowed by the time it takes to reject the middle item.

Clearly, an eye-tracking study would be needed to be sure, but it appears that the middle-color highlighting seems to draw the user's attention from where it might otherwise have been directed. If so, using color to emphasize and item in the middle of a menu might be useful to a designer who wants to draw attention to something important.

The serial-position effect for the top-color condition is harder to explain. The time to select the highlighted top item was about the same as selecting the top item under control conditions. But, rejecting the top item and selecting the middle item took slightly longer. Perhaps, as when color was used in the middle, subjects rejecting the highlighted item spent more time ensuring that the seemingly important item was not the one they wanted. But if this were the case, the time to select the bottom item should also be longer than controls. In fact, that time is shorter. The quadratic component of trend for the top color condition was not significant, so the shape cannot be relied on. Perhaps the drop in time for the bottom condition was incidental.

6. GENERAL DISCUSSION

Banner blindness occurs. Experiments 1 and 2 showed that users who search the web fail to notice items that are meant to be extremely obvious. This very "obviousness" ironically leads to decreased attention being paid those items. Banner blindness occurred with advertisements, non-advertisements that looked like advertisements, and plain text that was not ad-like. The fact that blindness occurred for graphical objects and small and large text objects indicates that features of the object itself are probably not the cause of the blindness but the relationship between the ignored object and other parts of the page somehow leads to banner blindness.

Experiment 1 showed that subjects were less likely to miss banners nearer to the main menu than banners at the very top of the page. And, it showed that users found banners more often when they were located at the bottom of the page hierarchy. By the time the subjects reached the bottom of the hierarchy they had exhausted their choices and were forced to look elsewhere or give up their search. Both of these findings indicate that subjects

focus their attention on the link-rich area of the page, the menu. Searching users do not start at the top left part of the page and skim it in its entirety, but instead visually pre-process the page for the area most likely to contain information important to their goal.

Experiment 2 explored ways of drawing more attention to banners while maintaining the "obviousness" of the banners by placing them near the top of the page. In the non-advertising portion of the experiment, only one group of subjects, those shown graphical banners perceptually grouped with the menu, showed evidence that a banner grabbed their attention. Facilitation did not occur for users exposed to unadvertisement-like text banners. Therefore, eliminating a resemblance to advertising did not make banner blindness disappear. And, increasing the perceptual grouping between the menu and the banner helped only a little. Finally, in the advertising portion of the experiment it was shown that animation did not draw the user's attention to the banners either.

Some recent studies have shown that banner advertisements are effective. One study showed that people who were exposed to an advertising banner and volunteered to participate in a survey were influenced by the content of the ad (Briggs & Hollis, 1997). An article in the New York Times indicated similar results. This article reported three findings: 1) sites that advertise more have more traffic, 2) spikes in site traffic coincide with banner advertising campaigns, and 3) users who are exposed to banner ads increase their brand awareness as demonstrated in subsequent surveys (Tedeschi, 1998).

The findings of Experiments 1 and 2 don't necessarily contradict these reports. First, Experiments 1 and 2 focused exclusively on users who were searching for facts. The findings described by Briggs and Hollis and by

Tedeschi either combine searching users together with browsing users (usage spikes during ad campaigns) or focus on browsing users only (people willing to fill out surveys). Secondly, Experiments 1 and 2 do not show that users never see or use banners, just that they are less likely to see and use banners than they are to use simple text links. Finally, Experiment 2 goes beyond advertising and examines attempts to attract attention to important or popular items through salient links created through non-graphical means like large and small text.

Web designers make certain items obvious because they want to ensure these items are noticed. The small amounts of mitigation achieved through changing grouping and formats of the banner in Experiment 2 certainly did not reach the level of "grabbing" the user's attention. Nearly all of the user's attention was apparently directed toward the menu. Therefore, Experiments 3 and 4 focused on how to attract a user's attention within a menu.

Within a menu, three characteristics were examined for effects on item selection and latency: text-size emphasis, color emphasis, and position. Text-size emphasis was not successful in positively influencing user behavior. Very large text used in Experiment 3 created some banner blindness, although to a much lesser degree than was found in off-menu banners in Experiments 1 and 2. Subtly larger text used in Experiment 4 did not created banner blindness, but it didn't reach the second goal of attracting attention to itself by affecting user behavior. It is possible that positive effects might occur at some medium level of this variable, but given the results found so far, it doesn't seem like a promising direction for further investigation.

Color emphasis achieved the dual goals of (1) not causing banner blindness and of (2) influencing user behavior to favor the colored item. The effect was statistically significant at both the top and the middle of the menu

in Experiment 3, when the subjects were selecting between two identical items. In Experiment 4, which was somewhat more realistic, the effect was only statistically significant when color was used at the top of the menu. Therefore, although position alone had little effect on choices made by control subjects, position in combination with highlighting did seem to have an effect.

6.1. Generalizability of On-Menu Highlighting

In considering the generalizability of the results of Experiments 3 and 4, it is important to note that there may be some interaction between the link content and potential serial position effects on selection. First, on control trials in Experiment 3, subjects chose the top of two possible items 85% of the time while control subjects in Experiment 4 chose from three possible positions at chance levels. One possible contributing factor to this might be that in Experiment 3 a statement above the menu indicated that the results shown were "in order of relevance." This might have influenced the subjects to favor items near the top. However, it is likely that many subjects did not notice the notation about the order. It was, after all, located outside the menu. The strong position effect in Experiment 3 probably occurred because the two choices were identical in content, and the subjects judged, through past experience or by reading the order notation on the page, that the one higher on the menu might be better. They might have used the heuristic, "all else being equal, choose the top one." When one item was highlighted with color, all wasn't equal. Therefore, the subjects chose the highlighted item at the top 100% of the time (compared with 85% of the time for controls) and they chose the lower highlighted item 64% of the time (compared with 15% for controls).

For control subjects in Experiment 4, the choices weren't equal. The three "good" options differed substantially, and subjects could make a value

judgment for which was best. Therefore, control subjects didn't need to resort to their secondary heuristic of using position to make a selection among the three good choices. Subjects in the experimental group could be influenced by that secondary heuristic if it was strong enough. The combination of position and color highlighting could make a pretty good choice overthrow a slightly better choice, and subjects reliably chose the top item with color highlighting more often than that same item was selected without the color highlighting. If this model is accurate, using emphasis in menus should influence behavior most when it emphasizes an item that is equally good or slightly worse than other items available on the menu.

Another consideration in generalizing these data is that the subjects in the study had only one opportunity to make a selection. Most web users have additional opportunities to make selections if the first one doesn't prove useful because they can return and make a different selection. Subjects in Experiments 3 and 4 could make only one selection. Therefore, they may have spent more time considering their decision, and the effect of position may have had less of an effect than it would have in the real world. Users in the real world may begin at the top of the list and select the first option that seems reasonable, returning to the list multiple times to choose each reasonable option in turn. Greater position effects would occur partly because the user would make selection from the top down in order to ensure an exhaustive use of the options. However, when a cost is involved in searching each item on the list, perhaps in page-load time or in effort in learning several different web-site structures, users may more carefully consider each option before making a choice. In this case the experimental results are more realistic and order position effects would probably be reduced.

Several aspects of banner blindness and item-emphasis in menus are still to be investigated. I did not look at the effect of color highlighting in menus over time. If users discover with experience that emphasized items are frequently equal to or worse than the other choices, the positive effect of highlighting might disappear or become negative.

Another issue is what effect known-to-be-popular links might have. If most users know a particular service to be popular, they might expect that link to be distinctive. For instance, a link to download browser software from Microsoft or Netscape might be expected to be distinctive. In this case, off-menu "banner" links might not cause banner blindness and in fact may grab users' attention.

6.2. Design Recommendations

One item separated visually from everything else on a web page may be completely ignored by web searchers, even by searchers who are deliberately searching for the information provided in that item. Designers should be cautious about following guidelines stating that increasing the visual distinction between "important" items and other items is desirable; the visual distinctiveness may actually make important items seem unimportant. Any item that is visually distinguished on a page should be duplicated within a collection of links or other navigation areas of the page. That way, if searchers ignore the large, salient item, they can still find what they need through basic navigation.

Any method that is created to make something stand out should be carefully tested with users who are specifically looking for that content to ensure that it does not cause banner blindness. The only method of attracting attention found here was using color to highlight something in a menu. Other strategies like bold, medium-sized text, or icons might also make and

item stand out in a menu. Off-menu distinctiveness might help if the item is one that users expect to be distinctive. The key is to test the method to ensure it works as the designer hopes.

REFERENCES

Aaltonen, A., Hyrskykari, A., & Räihä, K.-J. (1998). 101 spots, or how do users read menus? *Proceedings of ACM CHI'98 Conference on Human Factors in Computing Systems*, **1**, 132-139.

Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception & Psychophysics*, **19**(4), 361-367.

Beck, J. (1972). Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology*, **85**(1), 1-19.

Bednall, E. S. (1992). The effect of screen format on visual list search. *Ergonomics*, **35**(4), 369-383.

Briggs, R., & Hollis, N. (1997). Advertising on the Web: Is there response before click-through? *Journal of Advertising Research*, **37**(2), 33-45.

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. *Proceedings of ACM CHI'99 Conference on Human Factors in Computing Systems*, in press.

Compaq Computer Corporation. (1999). *AltaVista Search Service* [On-line]. Available: http://www.altavista.com

Cropper, A. G., & Evans, S. J. W. (1968). Ergonomics and computer display design. *The Computer Bulletin*, **12**(3), 94-98.

Detweiler, M. C., & Omanson, R. C. (1996). *Ameritech Web Page User Interface and Design Guidelines* [On-line]. Available: http://www.ameritech.com/corporate/testtown/library/standard/web_guidelines/index.html

DoubleClick Inc. (1997). *DoubleClick* [On-line]. Available: http://www.doubleclick.net

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**(3), 433-458.

Fisher, D. L., Coury, B. G., Tengs, T. O., & Duffy, S. A. (1989). Minimizing the time to search visual displays: The role of highlighting. *Human Factors*, **31**(2), 167-182.

Fisher, D. L., & Tan, K. C. (1989). Visual displays: The highlighting paradox. *Human Factors*, **31**(1), 17-30.

Fleming, J. (1997). *In Defense of Web Graphics: Graphic Designers Offer More Than Just Flashy Graphics* [On-line]. Available: http://www.webreview.com/97/07/25/feature/index4.html

Gillam, B. (1972). Perceived common rotary motion of ambiguous stimuli as a criterion of perceptual grouping. *Perception & Psychophysics*, **11**(1B), 99-101.

Gillam, B. (1981). Separation relative to length determines the organization of two lines into a unit. *Journal of Experimental Psychology: Human Perception & Performance*, **7**(4), 884-889.

Gomberg, B. F. (1985). *Strategic and Attentional Factors of Computer-Generated Highlighting*. Unpublished Master's Thesis, New Mexico State University, Las Cruces, NM.

Hornof, A. J., & Kieras, D. (1997). Cognitive modeling reveals menu search is both random and systematic. *Proceedings of ACM CHI'97 Conference on Human Factors in Computing Systems*, 107-114.

I/PRO Research/DoubleClick. (1996). *Web in Perspective: Ad Response* : Abstract from http://www.doubleclick.net/frames/siteiset.htm.

Kahneman, D., & Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In S. Dornic (Ed.), *Attention and Performance VI*, (pp. 307-332). Hillsdale, NJ: Erlbaum.

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, **16**(2), 177-216.

King, D. L. (1990). Gestalts are more closely associated with performance on a discrimination task than are component stimuli. *American Journal of Psychology*, **103**(1), 37-52.

Lee, E., & MacGregor, J. (1985). Minimizing user search time in menu retrieval systems. *Human Factors*, **27**(2), 157-162.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Pacific Grove, CA: Brooks/Cole.

McLaughlin, J. P. (1968). Recall and recognition measures of the von restorff effect in serial learning. *Journal of Experimental Psychology*, **78**(1), 99-102.

Morkes, J., & Nielsen, J. (1997). *Concise, SCANNABLE, and Objective: How to Write for the Web* [On-line]. Available: http://www.useit.com/papers/webwriting/writing.html

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nielsen, J. (1994). *Report from a 1994 Web Usability Study* [On-line]. Available: http://www.useit.com/papers/1994_web_usability_report.html

Nielsen, J. (1997). *Interface Design for Sun's WWW Site* [On-line]. Available: http://www.sun.com:80/sun-on-net/uidesign/

Norman, K. L. (1990). *The Psychology of Menu Selection: Designing Cognitive Control of the Human/Computer Interface.* Norwood, NJ: Ablex Publishing Corp.

Nygren, E. (1996, Oct). *"Between the clicks" Skilled users scanning of pages.* Paper presented at Designing for the Web: Empirical Studies. Available: http://microsoft.com/usability/nygrem.rtf.

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology,* **24**(3), 436-447.

Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception & Psychophysics,* **18**(5), 355-361.

Pulat, B. M., & Nwankwo, H. H. (1987). Formatting Alphanumeric CRT Displays. *International Journal of Man-Machine Studies,* **26**(5), 567-580.

Salon Internet Inc. (1999). *Salon Magazine* [On-line]. Available: http://www.salonmagazine.com

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition,* **19**(6), 523-542.

Shipley, T. F., & Kellman, P. J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception & Psychophysics,* **52**(1), 97-106.

Shneiderman, B. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (Second Edition).* Reading, MA: Addison-Wesley Publishing Co.

Smith, S. L., & Goodwin, N. C. (1971). Blink coding for information display. *Human Factors,* **13**(3), 283-290.

Spool, J. M., Scanlon, T., Schroeder, W., Snyder, C., & DeAngelo, T. (1997). *Web Site Usability: A Designer's Guide.* Andover, MA: User Interface Engineering.

Streveler, D. J., & Wasserman, A. I. (1984). Quantitative Measures of the Spatial Properties of Screen Designs. *Proceedings of IFIP INTERACT'84: Human-Computer Interaction,* 81-89.

Tan, K. C., & Fisher, D. L. (1987). Highlighting and search strategy considerations in computer-generated displays. *Proceedings of the Human Factors Society 31st Annual Meeting*, 524-528.

Tedeschi, B. (1998, Oct. 7). Study finds ad banners make an impression. *The New York Times*.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception & Performance*, **8**(2), 194-214.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **17**(3), 652-676.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance*, **16**(3), 459-478.

Tullis, T. S. (1983). The formatting of alphanumeric displays: A review and analysis. *Human Factors*, **25**(6), 657-682.

Tullis, T. S. (1988). A system for evaluating screen formats: Research and application. In H. R. Hartson & D. Hix (Eds.), *Advances in Human-Computer Interaction, Vol. 2*, (pp. 214-286): Ablex Publishing Corp.

Ullman, S. (1984). Visual routines. *Cognition*, **18**, 97-159.

Van Dam, G., Peeck, J., Brinkerink, M., & Gorter, U. (1974). The isolation effect in free recall and recognition. *American Journal of Psychology*, **87**(3), 497-504.

van Nes, F. L., Juola, J. F., & Moonen, R. J. A. M. (1987). Facilitation and inhibition of visual display search processes through use of colour. *IPO Annual Progress Report, Rpt No 22*, 117-123.

Williams, J. R., & Leaf, W. A. (1995). Subjective and objective judgments of screen formats. *Proceedings of ACM CHI'86 Conference on Human Factors in Computing Systems*, 689-693.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**(2), 202-238.

Woodward, R. M. (1972). Proximity and direction of arrangement in numeric display. *Human Factors*, **14**, 337-343.

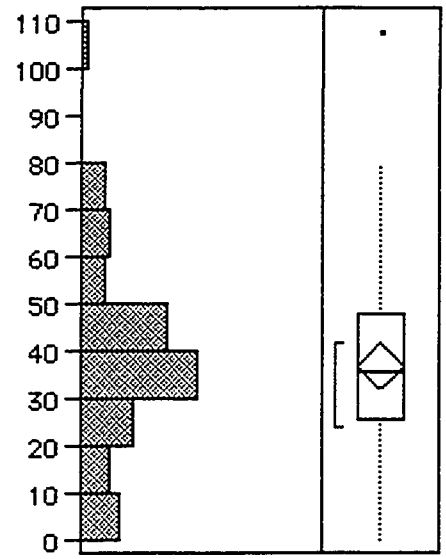
Yahoo! Inc. (1997). *Yahoo!* [On-line]. Available: http://www.yahoo.com

APPENDIX A: QUESTIONNAIRE RESULTS FOR EXPERIMENT 2

The questions and summaries of results of the post-experiment questionnaire used in Experiment 2 are shown below. The original text of the items is given in bold, with the results following.

1. Did you read and follow all instructions in this experiment?
72 of 72 (100%) respondents answered "yes".

2. How long has it been since you used the Web for the first time?
Fields for years and months were given. The distribution of responses in months is shown here:



The mean was 37.2 month experience, with a standard deviation of 20.2. It is very likely that the phrasing of this question or respondents' faulty memories make the responses to this item invalid. Few people had any web access before January 1994. Therefore, since this experiment took place in March 1998, the maximum number of months experience should only vary rarely exceed 51. The relatively large number of subjects who claimed near and over that number makes this item suspect, particularly since most of the experiment participants would have been in junior high or early high school in January 1994.

3. How often do you use the Web?
Every day - 31 (43%)
Several times a week - 34 (47%)
Several times a month - 5 (7%)
Once a month - 1 (1%)
Less than once a month - 1 (1%)

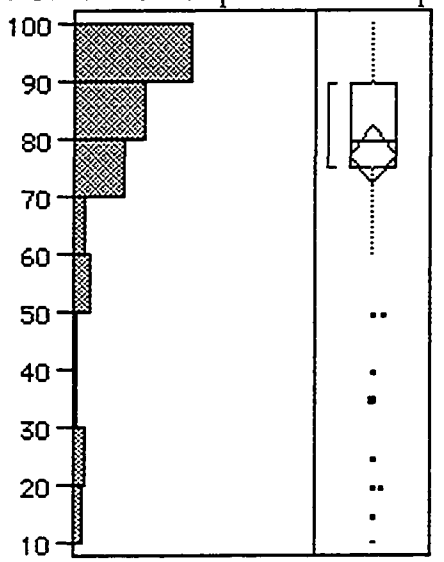
4. Think about how much you use the web to look for specific information vs. how much time you spend browsing, or not looking for anything specific.

Looking for specific information could include trying to find specific people, phone numbers, today's weather, airline flight schedules/reservations, or other facts.

Browsing could include surfing around for "something interesting." reading up on the news of the day, or other situations where you aren't trying to answer specific questions.

Subjects were forced to respond with the percent of time they spend browsing and searching with the constraint that these add to 100%.

Distribution of percent time spent searching:



The mean percentage time spent searching was 77% with a standard deviation of 22%.

5. Have you ever used the Web to look for information on products or services you were interested in purchasing?

49 of 71 (69%) respondents answered yes.

6. Have you ever bought anything using the Web?

26 of 71 (37%) respondents answered yes.

7. Have you ever given a credit card number over the Web?

21 of 71 (30%) respondents answered yes.

8. To what extent do you find advertising on the Web to be useful?

Never useful - 17 (24%)
Occasionally useful - 37 (52%)
Sometimes useful - 11 (15%)
Often useful - 4 (6%)
Extremely useful - 2 (3%)

9. To what extent do you find advertising on the Web to be annoying?
Never annoying - 2 (3%)
Occasionally annoying - 11 (15%)
Sometimes annoying - 19 (27%)
Often annoying - 27 (38%)
Extremely annoying - 12 (17%)

10. What do you think about the amount of commercialism on the Web?
There should be more - 1 (1%)
It's about right - 34 (48%)
There's too much - 36 (51%)

11. Think about animations like this one: [animated folding letter]

11a. To what extent do you find animations to be useful or entertaining?
Never useful/entertaining - 2 (3%)
Occasionally useful/entertaining - 15 (21%)
Sometimes useful/entertaining - 22 (31%)
Often useful/entertaining - 26 (37%)
Extremely useful/entertaining - 6 (8%)

11b. To what extent are animations annoying?
Never annoying - 7 (10%)
Occasionally annoying - 34 (48%)
Sometimes annoying - 22 (31%)
Often annoying - 8 (11%)
Extremely annoying - 0 (0%)

12. In today's experiment, did you notice the ads that were shown to you during the search tasks you performed?

14 of 71 (20%) respondents answered yes.

13. In today's experiment, did you read the advertisements shown to you during the search tasks?
No, none of them - 55 (77%)
Yes, a few of them - 15 (21%)
Yes, most of them - 1 (1%)

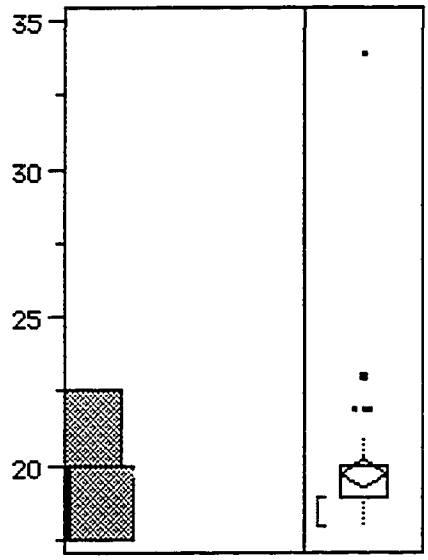
14a. In today's experiment, do you recall seeing text or a graphic that helped you with your task? For example, that told you that the "Shoe Rack" you were looking for was in item C5? (The actual item mentioned would have been different - not the shoe rack.)

14 of 71 (20%) respondents answered yes.

14b. If Yes, did you use that information to help you find the item more quickly?
No - 2 (3%) [Both of these respondents answered no to #14a.]
Didn't see them - 55 (76%)
Yes - 14 (20%)

15. In today's experiment do you recall seeing text or a graphic that told you the location of an item you were not looking for? For example, it might have told you where the shoe rack was when you were looking for the book about opera.
11 of 71 (15%) respondents answered yes.

16. What is your age?
Distribution:



Mean age was 19.8 years, with a standard deviation of 2.

17. What is your gender?
39 of 71 (55%) respondents answered female.

APPENDIX B: ANOVA RESULTS FOR EXPERIMENT 2, NON-ADVERTISING BLOCK

The mixed-factor analysis of variance results for the non-advertising-trials block in Experiment 2 are shown below. For all repeated-measures analyses of variance reported here, a multivariate approach was used. All repeated measures use the F test of the Wilkes' Lambda test statistic (Maxwell & Delaney, 1990).

Variables

Trial Type: Within-subjects, three levels. Relevant banner (experimental trial), irrelevant banner (control), and unhelpful banner (control).

Format: Between subjects, two levels. Graphical, large text, small text.

Grouping: Between subjects, two levels. Banner grouped with menu or title.

Block Order: Between subjects, two levels. The non-advertising block occurred before or after the advertising block.

Effect	Wilks' λ	F	Df	p
Trial Type	0.968	0.976	2, 59	0.383
Trial Type x Format	0.996	0.052	4, 118	0.995
Trial Type x Grouping	0.922	2.502	2, 59	0.091
Trial Type x Block Order	0.992	0.232	2, 59	0.793
Trial Type x Grouping x Format	0.939	0.940	4, 118	0.444
Trial Type x Grouping x Block Order	0.980	0.601	2, 59	0.552
Trial Type x Format x Block Order	0.928	1.121	4, 118	0.350
Trial Type x Grping x Format x Block Ord	0.996	0.057	4, 118	0.994

APPENDIX C: ANOVA RESULTS FOR EXPERIMENT 2, ADVERTISING BLOCK

The within-subjects analysis of variance results for the advertising trials block in Experiment 2 are shown below. A multivariate approach was used for the repeated-measures analyses. All repeated measures use the F test of the Wilkes' Lambda test statistic (Maxwell & Delaney, 1990).

Variables

Shown: Two levels. Advertisements that were shown vs. those that were not (controls).

Grouping: Two levels. Banner grouped with menu or title.

Animation: Two levels. Animated or still.

Context: Two levels. The advertisement was tested in it original graphical form (high context) or with a text representation of the brand name (low context).

Effect	Wilks' λ	F	Df	p
Shown	0.899	7.971	1, 71	0.006
Shown x Grouping	0.998	0.111	1, 71	0.740
Shown x Animation	0.997	0.198	1, 71	0.658
Shown x Context	0.995	0.345	1, 71	0.559
Shown x Grouping x Animation	0.998	0.114	1, 71	0.737
Shown x Grouping x Context	0.988	0.868	1, 71	0.355
Shown x Animation x Context	0.986	0.998	1, 71	0.321
Shown x Grouping x Animation x Context	0.955	3.335	1, 71	0.072

APPENDIX D: TASKS USED IN EXPERIMENTS 3 AND 4

These are the tasks used in Experiments 3 and 4. The first eight tasks were used in Experiment 3. Task 9 was added in Experiment 4.

Task 1: You would like to know how well the stock market is doing. Find out what the Dow Jones average closed at yesterday.

Task 2: You followed Mark McGuire's home run record for most of the baseball season, but you are not sure what his batting average for the year was. Find out.

Task 3: Someone recommended the clothing catalog "Lands End" to you, and you'd like to order something. Find out the ordering phone number.

Task 4: You are doing a project on profiles of executives in American corporations. Find out the name of the CEO of Coca Cola.

Task 5: You are interested in taking a trip to DisneyWorld in Florida. Find out the price of admission for one day.

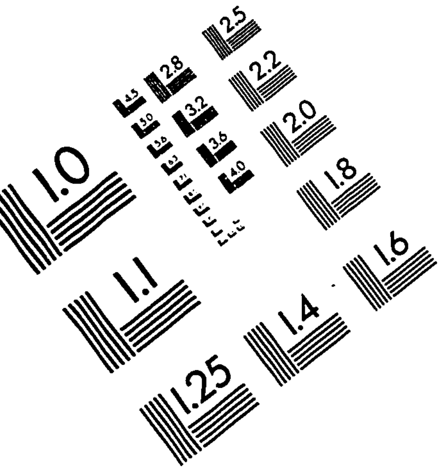
Task 6: Someone recently recommended a book to you, called *Peopleware*. But you don't know who wrote it. Find out the name of the author(s).

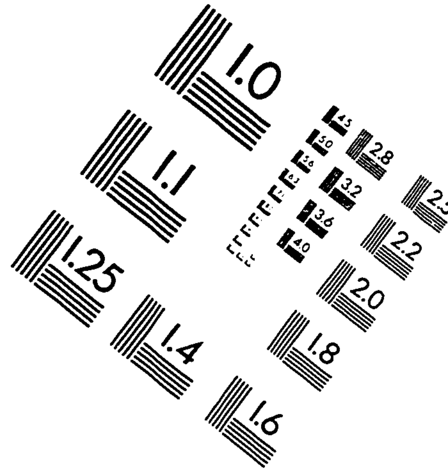
Task 7: You are going to a conference in Chicago soon, at the Downtown Marriot. Find out the phone number to call for reservations.

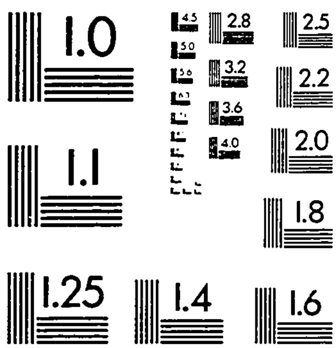
Task 8: You were thinking about African geography lately, and realized you can't remember the name of the capital of Uganda. Find out the name of the capital city.

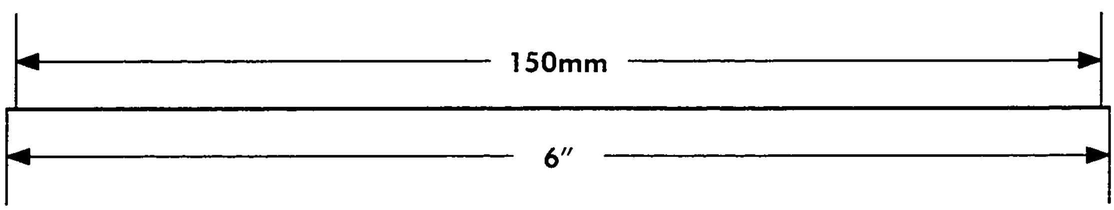
Task 9: After a recent bout of food poisoning, you are interested in food safety. Find some tips on how to prevent food poisoning.

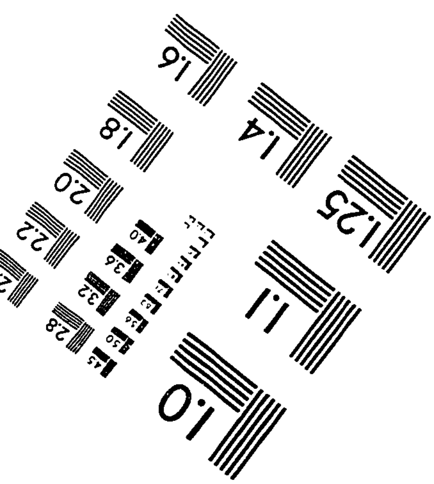
IMAGE EVALUATION TEST TARGET (QA-3)

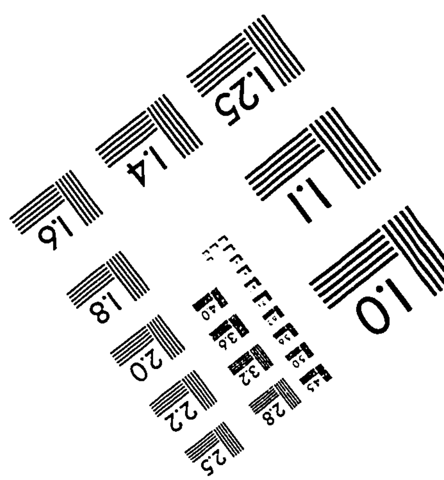












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989