The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task

Robert Pearson1, Paul van Schaik*

Psychology Section, School of Social Sciences and Law, University of Teesside, Middlesbrough TS1 3BA, UK

Received 7 November 2002; received in revised form 28 January 2003; accepted 16 February 2003

Abstract

This study aimed to investigate the validity of psychological experimental methods within human–computer interaction research (Carroll, 1989) and to examine design guidelines pertaining to hypertext link colour and positioning of navigation menu frames as part of web documents. The results of past research on both link colour and positioning of menus are mixed and guidelines are usually not based on empirical evidence (Tullis, 1997; Shneiderman, 1997). The study used a repeated measures experimental design. Participants carried out both a visual search task and an interactive search task. Task performance on the two tasks did not to correlate ($p > 0.05$), indicating that the visual search task may lack external validity. Results of the interactive search task suggest that the design convention of blue links (Nielsen, 1999a) should be retained as responses for blue were found to be significantly quicker than red, $F(1,117) = 14.526, p < 0.001, MS_{\text{colour}} = 89.866$. Furthermore, an effect of presentation position, $F(3,117) = 8.410, p < 0.001, MS_{\text{position}} = 61.015$, was found, with support for menus on the left (Nielsen, 1999a; Campbell & Maglio, 1999) or right (Nielsen, 1999a). Evidence was also found to support the conjecture that experienced Internet users might have formed automatic attention responses to specific web page designs. The need for validation of behavioural and psychometric methods with task performance and the use of cognitive-perceptual-motor modelling are discussed.

© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: World Wide Web; Frame layout; Link colour; Task performance; Visual search; Interactive search; Automatic attention response; Experimentation

*Corresponding author. Tel.: +44-1642-342320; fax: +44-1642-342399.
E-mail address: p.van-schaik@tees.ac.uk (P. van Schaik).
1Now at Nissan Motor Manufacturing (UK) Limited, Sunderland, UK.
1. Introduction

The number of World Wide Web (Web) sites is growing at an astronomical rate. The Netcraft Web Server Survey (http://www.netcraft.com) in July 1995 identified 19 000, in April 1997, 1 000 000, in March 2000, 13 000 000 and in October 2002, 38 000 000. This does not include intranets so the actual total web sites (intranet and Web) will be larger. Because of this uncontrolled growth of the Web and the absence or disregard of a good and comprehensive set of web interface design guidelines, millions of poorly designed Web pages crowd the Web (Vora, 1997).

A key factor in retaining visitors to web sites is the usability of these sites (Nielsen, 2000). A particular factor that affects usability is screen design. Poor screen design can also have a detrimental effect on a user’s performance by decreasing search speed and provoking errors (Streveler and Wasserman, 1984).

Design guidelines can help web designers to produce more usable designs if they are based on empirical research (van Schaik and Ling, 2001a, b). However, currently many guidelines are still based on personal experience and observations (Streveler and Wasserman, 1984; Shneiderman, 1997; Borges et al., 1998) and therefore their validity is questionable.

In addition, the usability of sites can be enhanced by applying software usability engineering techniques that have been applied for more than a decade to software systems other than web sites. Unfortunately, there is a lack of evidence that these techniques are widely applied to the development of web sites (Mayhew, 1998).

1.1. Visual search

Information retrieval is a significant activity in using the Web. Because this activity invariably involves visual search, research on visual search processes is essential. Two basic stages are involved in visual search (Scott, 1993; Wolfe, 1994). At first, all screen locations are processed in parallel, but only a limited amount of information is extracted. Subsequently, a second stage restricts itself to areas of the visual field that are of interest. At this stage more complex tasks can be carried out and more information extracted. However, this stage has limited capacity and information can only be extracted from one or, perhaps, a few spatial locations at a time. Two visual sub-systems work together when viewing dynamic displays to produce a scan path. The peripheral sub-system carries out the initial parallel processing and determines where the next point of attentional fixation will be placed. The foveal sub-system is then directed to this point and performs more complex operations such as reading and object identification. Visual conspicuity is a factor that influences the selection process executed by the peripheral system. This conspicuity can be defined as a combination of an object’s properties, relative to its background, which attracts attention and is therefore seen and is based on colour, shape, size, orientation and other properties of the display. According to Nygren (1996) search time can be up to 83% faster if a target object is made visually conspicuous.
1.2. Colour in screen design

An example of the application of colour in displays is the use of blue as the default colour of hypertext links on web pages. Nielsen (1999a) claims that “the mother of bad design conventions is the decision to make hypertext links blue”. The reason for this is that a sharp blue image is impossible to obtain, as the wavelength of light creating blue can never be brought into focus. This occurs for people of all ages but as people get older their sensitivity to blue decreases because it becomes harder to focus with age (Galitz, 1997). Also only 2% of the cones in the eye—the “blue-sensitive” cones—react to short wavelengths. The very centre of the retina is devoid of these cones and consequently creates a “blue-blindness” for small objects fixated upon it (Murch, 1984).

Of the cones, 64% are especially sensitive to red (Murch, 1984), but there are very few of these in the periphery so red cannot be clearly seen there. Blue on the other hand has been found to be easily detected in the periphery due to an increased number of cones towards the edge of the retina. From the previous discussion it follows that blue-sensitive cones will be particular useful for the first stage of visual search and red-sensitive cones for the second stage.

These findings have led to guidelines stating that blue should not be used for text or small objects but is good for a background colour (Murch, 1984; Galitz, 1997; Marcus, 1997; Tullis, 1997). According to Galitz (1997), warm colours such as, red, yellow and orange, imply active situations or actions that are necessary and so force a user to attend to them. On the other hand, cool colours such as, green, blue, violet and purple, imply background or status information.

A significant shortcoming of these guidelines though is that they do not take the background/foreground colour combination of a particular screen into account (Woods et al., 1992).

Some research seems to contradict these guidelines as blue has been found to be a good text colour when white is used as the background colour (Murch, 1984; Galitz, 1997; Marcus, 1997; Tullis, 1997). According to Galitz (1997), warm colours such as, red, yellow and orange, imply active situations or actions that are necessary and so force a user to attend to them. On the other hand, cool colours such as, green, blue, violet and purple, imply background or status information.

According to Galitz (1997) blue has been found to be acceptable by some researchers and unacceptable by others because exact measures of saturation and hue being rarely given, thereby confounding research results. The use of colour names has also been claimed to underspecify the visual stimulus. For example Woods et al. (1992) state that blue can run from a deeply saturated blue to a pale cyan and it is the saturated blue that is difficult to read when an object is relatively small or thin.

1.3. Frame layout

In relation to frame layout, various authors have suggested that screens should be structured in an organized manner to improve visual search and therefore usability
(Graf and Krueger, 1989; Marcus, 1992, 1997; Galitz, 1993, 1997; Comber and Maltby, 1997; Tullis, 1997). According to Marcus (1997) there are three primary means of achieving an organized screen layout: (1) use an underlying grid layout, (2) standardize the screen layout and (3) group related items. Designers who employ frames attempt to do all of the above, so the implementation of frames within a web document should have a positive effect on usability. In addition, the support of frames by web browsers may provide a partial solution to the problem of navigation; this is because frames permit the laying out of the web browser window into sections, which users can view and interact with individually (Vora, 1997). This allows web designers to provide navigation controls in one frame that always stay visible, even if the contents of another frame change (Bain and Gray, 1997; Hobbs, 1999). Although there is research into graphical user interfaces and menus (for example Tullis, 1997) that could be generalized to and applied to the use of frames for navigational purpose in a web page, the research appears to be inconclusive in helping to form design guidelines on whether the navigation frame of a web page should be placed vertically (either on the left or right) or horizontally (either on the top or bottom).

Nielsen (1999a) highlights the strong convention of Web pages having a coloured stripe down the left side of the page containing the main navigation links, which he terms “Yellow Fever”. This idea is supported by Campbell and Maglio (1999) who found that stimuli positioned in the upper left of the screen were the quickest to be identified and lower right the slowest. Galitz (1997) also states that eye fixation studies have shown that people first look at the upper left of a display and then quickly move through the display in a clockwise direction. It has also been suggested though that navigation should be placed at the top and bottom of pages that require scrolling (Borges et al., 1998). Another eye tracking study, however, has found that users ignore the last \( \frac{1}{3} \) of a Web page (Spool, 1999). This may mean that important information, such as navigation links, should not be placed at the bottom of a screen.

Nonetheless, Nielsen (1999a) criticizes Yellow Fever for a number of reasons. Firstly, it takes a large amount of the screen (20%), even after the user has scrolled to the bottom of the page. This scrolling also causes the links to vanish (a problem not associated with frames; Bain and Gray, 1997; Hobbs, 1999). This means that if a screen is long users have to scroll back up the screen to find the relevant links before proceeding.

Secondly, the placing of the navigation links on the right would be better as they would be closer to the scroll bar and therefore abide by Fitts’ law (Nielsen, 1999a). Fitts’ law states that the time taken to acquire a target (by moving the cursor) is a function of the distance that needs to be travelled and the size of the target (Tognazzini, 1992). Thus, having the navigation rail (frame or stripe) on the right side of the screen next to the scroll bar should save a users time if scrolling is required.

Another reason for placing the navigation on the right is that it has been claimed that users “always look at the content first” when they first access a new Web page (Nielsen, 1999a). Since in Western cultures people read from left to right, this would therefore mean that the content of the page should be placed on the left of the screen.
and the navigation rail on the right as once a user had finished viewing the content their gaze would naturally shift to the right to decide where to go next.

1.4. The role of previous knowledge

Even though Nielsen (1999a) argues that having blue as the default link colour and the convention of navigation links being placed on the left are wrong, he is also of the opinion that it would be wrong to change them even if a change in principle would improve usability. This is due to users knowing, Nielsen (1999a) claims, what blue text means, that is “click here”, on the Web. He states that the time saved in knowing what to do on a page is probably much greater than the time saved by having the word written in a colour that is quicker to read. This idea is supported by research that has found that detection can become automatic if a user consistently searches the same environment for the same information and this information is always represented in the same way. In this way an automatic attention response is formed (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). Marks and Dulaney (1998) claim that for Web users an automatic attention response has formed for links, as users automatically process underlined text in a different colour as a link. This idea is supported by the finding that search time is quicker for targets in the form of links (blue underlined) than normal black text on a white background (Campbell and Maglio, 1999), even though the black should be in principle easier to read. Related to explanations in terms of automatic attention response, but based on a different theoretical perspective, the phenomenon of task facilitation with familiar interface styles and task degradation with unfamiliar styles has also been explained by users’ schemas of web pages in terms of their appearance and user–system interaction (Norman, 1988, 1999).

In terms of the position of the navigation links, again Nielsen (1999a) raises his point that although changing the standard at this time may improve the usability of a site by a small amount, it would impose a much greater cost due to confusion and reduced ability to navigate smoothly.

This is due to the fact that previous knowledge can influence effective search of a visual display (Marks and Dulaney, 1998). If users know where to focus the search then they will be quicker to locate the relevant information (Posner et al., 1978), as they will be carrying out a “direct search” (Scott, 1993). If information appears in an unexpected position however search time will increase (Posner et al., 1978). So again web users may have formed an automatic attention response in terms of where relevant information will be found for navigation.

If this is the case then, for experienced web users, keeping the default link colour blue and placing the navigation frame on the left of the screen may be advantageous. If this is not the case and also for inexperienced users, these factors need to be investigated to discover what makes the most usable screen. This is because, even if time savings from improving the design of a web page are short, due to the frequency of tasks time savings can be considerable (Tullis, 1981) and due to users changing their task strategy with small changes (in the order of tens of milliseconds) in system response times (Gray and Boehm-Davis, 2000).
1.5. Validity of experimental research

As discussed previously, current guidelines and standard design procedures for the design of graphical user interfaces and web pages are not conclusively backed by empirical evidence. One reason for this is due to the conflicting results obtained from experimental studies. This also suggests that this area needs further investigation so that accurate design guidelines can be produced. This is where a problem lies, as there is a debate on the role of psychological experimentation in human–computer interaction research. Whiteside and Wixon (1987), Carroll (1989) and Winograd and Flores (1986) hold a strong view that controlled psychological experiments do not have a place in human–computer interaction research.

Carroll (1989, p. 53) refers to psychological experiments within human–computer interaction as “toy problems” and states that “there is no way to know whether a toy problem is representative of a real problem”. He also states that direct contrast laboratory work, in terms of comparing two alternative designs, entails compromises in the face validity of the work that leads to a failure to produce definitive evaluations. Carroll also claims that direct contrasts, in terms of psychological experiments, cannot evaluate something that does not yet exist, so experimental research always lags development by some fraction of a development cycle.

This perspective leads onto the idea that the only way to test a Web site’s usability is to observe real users carrying out real tasks (Bevan and Mcleod, 1994; Spool et al., 1997; Nielsen, 1999b; Diaper and Waelend, 2000). Comber and Maltby (1997) argue that this means that it is important for the design being used for testing to have “real utility”, that is one that could be perceived by the participants as being a real application rather than a contrived experimental one. In the case of web page design this is simplified by the fact that the product being tested can be easily written in the same programming language that an actual web page would be written in, for example HTML or Java, and so has real utility.

Other researchers hold the view that psychological experimentation can be relevant for the development of both human–computer interaction theory and practice (Card et al., 1983; Newell and Card, 1985; Barnard, 1991; John and Vera, 1992; Zhang, 1999; Payne et al., 2001). In this context, Payne et al. have defined the paradigm of “interactive search” as a situation where “a task is accomplished as a sequence of selections from displayed sets of alternatives”. It is precisely this more realistic type of task that has recently been used in experimental psychological research (for example Payne et al., 2000).

Considering the different views regarding the type of psychological research that is required in human–computer interaction in general and for designing web-based applications in particular, this study included two experimental tasks. The first was a visual search task using the choice reaction paradigm (van Schaik and Ling, 2001b). The second was an interactive search task (Payne et al., 2001), with participants carrying out information retrieval tasks as they would on the Web. The results from these two tasks were then analysed to see if they correlated. One aim of this was to investigate the external validity of the visual search task. Another aim was to contribute, based on experimental evidence, to design guidelines and conventions.
pertaining to the use of hypertext link colour and positioning of frames as a navigational device for web pages.

1.6. Hypotheses

This study investigates the effect of two independent variables, presentation position of hypertext links and link colour, on visual search performance and on interactive search task performance.

The following experimental hypotheses are tested: (1) mean performance differs between four presentation positions, (2) mean performance differs between two link colours and (3) the pattern of mean performance for the independent variable position differs between the levels of independent variable colour and vice versa, indicating a lack of independence between the two design parameters presentation position and link colour.

2. Method

2.1. Design

Two tasks—a visual search task and an interactive search task—were included each using an experimental $4 \times 2$ repeated measures design. In both cases the independent variables were presentation position, which had four levels, right, left, top, and bottom, and link colour, which had two levels, blue and red. For the visual search task the dependent variables were: percentage of hits, reaction time for hits, percentage of correct rejections and reaction time for correct rejections. For the interactive search task the dependent variables were time-on-task for correct answers and percentage of errors. In the visual search task data were recorded by computer; all the presentation of stimuli was randomized differently for each participant by computer. In the interactive search task data were recorded by hand using a video camera and stopwatch; a $32 \times 32$ Latin Square was produced, randomizing the questions to counteract order effects.

2.2. Participants

There were 40 participants, all university students, taking part in the research voluntarily. They consisted of 29 females and 11 males, with an age range of 19–49 years old and a mean of 25.8. All participants had 20/20 or correct vision and none had any colour vision deficiencies. In order to take part in the study, participants had to have at least some level of computer literacy and have experience with the Web or an intranet site. Only four participants had used the Web for less than a year and only one had used it for less than 3 months. The majority of participants (31) used the Web more than once a week and only one used it less than once a week (see Table 1). Therefore the majority of the participants were experienced Internet users.
2.3. Apparatus and materials

A computer with the following specifications was used to run both tasks in the experiment: Intel Pentium II 400 MHz, 64 megabyte RAM PC running Microsoft Windows 95, using a 3D Rage Pro AGP 2X graphics card, with a 14 in monitor, Windows 95 compatible keyboard and a Microsoft IntelliPoint compatible mouse (with trackball turned off).

The display properties of the graphics card were set at high colour (16bit), with screen dimensions of 800 × 600 pixels and contrast and brightness were both set at mid range. The experiment took place in an interview room, which was lit with overhead fluorescent lights of constant intensity typical of office conditions and was adequately sound proofed.

For both tasks the web pages were coded in HTML. These web pages consisted of two frames, a larger frame (82%, the main content) and a smaller frame (18%, the navigation menu section) containing five hypertext links. In both tasks the menu frame could be at the right, left, top or bottom and the menu links could be either blue (colour code #0000FF) or red (colour code #FF0000) depending on which level of the independent variables it represented. In the visual search task asterisks represented the main content (Fig. 1a) and in the interactive search task the main content was made up of information related to the question that had been set (Fig. 1b).

For the visual search task and its related practice, the stimuli were converted from web pages, displayed with Microsoft Explorer 4, into picture files (bmp format). This part of the experiment was then produced using SuperLab 1.05. In order to reduce the chance of word frequency effects occurring, all menu items (targets and non-targets) were nouns with a frequency range of 1 million to 66 million, with an average of 14.09 million (Johansson and Hofland, 1989). For the practice 16 stimuli were produced and for the main the visual search task another 128.

For the visual search task and its related practice, the stimuli were converted from web pages, displayed with Microsoft Explorer 4, into picture files (bmp format). This part of the experiment was then produced using SuperLab 1.05. In order to reduce the chance of word frequency effects occurring, all menu items (targets and non-targets) were nouns with a frequency range of 1 million to 66 million, with an average of 14.09 million (Johansson and Hofland, 1989). For the practice 16 stimuli were produced and for the main the visual search task another 128.

The web pages for the interactive search task and its related practice, were run with the use of Microsoft Explorer 4 and response times to questions were recorded with the use of a video camera and stopwatch. The mock web pages in this part of the experiment were made more externally valid by including both questions that required the screen to be scrolled to find information—and others that did not
require this—and questions that required participants to go through two levels of hypertext links before a correct answer could be found—and others requiring only one level of links. These variables were distributed evenly across the levels of the two independent variables. The 32 questions drew answers from eight web sites (in total 120 web pages), representing four domains: sport, shopping, music and computer equipment. The four questions for the practice were presented in the same order to each participant.

2.4. Procedure

Participants carried out both tasks individually.

2.4.1. Practice for the visual search task

Participants were first presented with a set of instructions. After reading the instructions participants had to press “s” to continue. A trial was then presented. This consisted of a blank white screen (shown for 2000 ms), a black target word on a white background (for 3000 ms, in Arial bold, size 48, in the centre of the screen) and a blank white screen (for 2000 ms); this was followed by a mock web page containing a menu frame (either at the right, left, top, or bottom) with 5 hypertext links (either in red or blue). The target word was either absent or present in the list of links and participants had to respond with either an “a” (absent) or a “p” (present) as quickly as possible; if a participant did not respond within 5000 ms an error was recorded and the next stimuli group was presented. The practice consisted of 16 trials (for a breakdown see Table 2). Once the practice was completed participants were asked if they had any questions or problems; if so these were discussed. They were then given the main visual search task.

2.4.2. Main visual search task

Participants were given the same instructions as in the practice; after reading them they had to press “s” to continue. They were then presented with the first trial and
had to respond in the same way as in the practice. After 64 trials participants were given a 1-min rest. The same instructions were then presented again and the experiment continued. Once participants had completed all 128 trials they were shown an end screen which thanked them for their participation in this part of the study. (See Table 2 for a breakdown of trials.) Participants were then given a 2-min rest, after which they were given the practice for the interactive search task.

2.4.3. Practice for the interactive search task

Participants were presented an instruction screen. Once they had read the instructions they clicked on the “continue” link to proceed. Participants were then presented with the first question. Once they had read it they clicked on a hypertext link in the centre of the screen to continue. Participants were then presented with a web page, in this the menu frame was either at the right, left, top, or bottom and the link colour was either blue or red. Participants then had to answer the question that had been set by clicking on a relevant link and reading the information provided. If participants believed they had followed a wrong link or had forgotten the question they could use the back button on the browser to move back. Once they had found what they believed to be the correct answer they clicked on the “answer” link. This took them to the answer screen containing a text box, into which they had to enter their answer. Once they had done this they clicked on the continue link to get the next question. Participants were also told that if they could not find what they believed to be the correct answer they should click on the answer icon and continue without typing an answer. The practice consisted of four questions and once these were completed participants were asked if they had any questions or problems; these were then discussed. They were then given the main interactive search task.

2.4.4. Main interactive search task

Participants were given the same instructions as in the practice. They then had to answer 32 questions in the same way they did in the practice. These 32 questions were presented in a random order for each participant with use of a Latin square. Participant 1 was presented order 1, participant 2 order 2 and so on, until participant 33 who was presented order 1, then participant 34 order 2 etc. The time taken for a
participant to answer a question was measured from when they clicked on continue after reading the question to clicking on answer to type their answer to the question. This was recorded through the use of a video camera and a stopwatch. If a participant answered a question incorrectly, or failed to answer it, this was recorded and the time for that question was discounted.

After the main interactive search task, participants’ demographics were then recorded. Participants were also asked for their preference pertaining to the position of the navigation menu frame (either right, left, top, or bottom) and the colour of the hypertext links (either blue or red). Participants were then debriefed by outlining the aims of the study. Overall, the whole experimental procedure took between 40 and 50 min to complete per participant.

3. Results

For the purposes of analysis of the visual search task, the data were divided into hits and correct rejections, which were each, examined in terms of accuracy and speed (that is reaction time). For the interactive search task, percentage of correct answers and time-on-task for correct answers were analysed. T tests indicated there was no difference between the performance of the most experienced participants (> 2 years experience) and those who had been using the Web for less than 6 months (all tests $p > 0.05$). For both tasks, analysis of variance (ANOVA) and tests of simple effects were used to assess the effect of presentation position and link colour in terms of differences in mean scores; post hoc analysis was conducted to test for differences in performance between layouts. Correlations were calculated between corresponding performance measures in the two tasks. Finally, participants’ preference for presentation position and link colour was analysed.

3.1. Results for the visual search task

3.1.1. Analysis of hits

3.1.1.1. Accuracy. A hit was defined as a correct response to a target word that was present in the navigation frame of a subsequently presented web page and the percentage of hits was calculated. Neither position, $F(3, 117) = 2.360, p > 0.05$, $MS_{\text{position}} = 280.43$, nor link colour had an effect on hits, $F < 1$. However, there was an interaction between position and link colour, $F(3, 117) = 6.977, p < 0.001$, $MS_{\text{interaction}} = 805.176$ (see Table 3 and Fig. 2). Simple effect tests showed that the effect of position was significant for blue links, $F(3, 117) = 3.706, p < 0.05$, $MS_{\text{position}} = 446.289$ and for red links, $F(3, 117) = 5.617, p < 0.005$, $MS_{\text{position}} = 639.323$. Further simple comparison tests showed top presentations were more accurate with red links ($p < 0.01$) and bottom presentations were more accurate with blue links ($p < 0.05$). Simple comparison tests with Sidak correction showed the bottom position was significantly more accurate than the top position ($p < 0.05$) with blue links, but the top position was significantly more accurate than the left position ($p < 0.01$) and the bottom position ($p = 0.01$) with red links.
3.1.1.2. Speed. Reaction times for hits were examined. The effect of position was significant, $F(3, 117) = 37.311$, $p < 0.001$, $MS_{\text{position}} = 547843.832$ (see Table 4). However, the effect of link colour, $F < 1$ and the interaction effect were not significant, $F(2.517, 98.159) = 2.847$ (Greenhouse-Geisser correction applied due to lack of sphericity), $p > 0.05$, $MS_{\text{interaction}} = 44676.599$. Multiple comparisons with Sidak correction showed that right and left positions did not differ and top and bottom positions did not differ. However, both the top and bottom positions were faster than both the right and left positions (all four $p < 0.001$).

3.1.2. Analysis of correct rejections
3.1.2.1. Accuracy. A correct rejection occurred when a participant correctly responded that a target word was absent in the navigation frame of a subsequently presented web page. Neither position, $F(3, 117) = 2.186$, $p > 0.05$, $MS_{\text{position}} = 59.082$, nor link colour, $F(1, 39) = 1.486$, $p > 0.05$, $MS_{\text{colour}} = 40.583$, nor the interaction effect, $F(3, 117) = 1.867$, $p > 0.05$, $MS_{\text{interaction}} = 60.384$, were significant and results appeared to show a ceiling effect (see Table 5).

3.1.2.2. Speed. Reaction times for correct rejections were examined. Link colour had no effect on performance, $F(1, 39) = 1.972$, $p > 0.05$, $MS_{\text{colour}} = 17779.540$.

### Table 3
Descriptive statistics for percentage of hits (visual search task)

<table>
<thead>
<tr>
<th>Link colour</th>
<th>Right</th>
<th></th>
<th>Left</th>
<th></th>
<th>Top</th>
<th></th>
<th>Bottom</th>
<th></th>
<th>Overall</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
</tr>
<tr>
<td>Blue</td>
<td>90.62</td>
<td>12.58</td>
<td>88.44</td>
<td>13.69</td>
<td>84.06</td>
<td>16.26</td>
<td>91.56</td>
<td>12.46</td>
<td>88.67</td>
<td>10.05</td>
</tr>
<tr>
<td>Red</td>
<td>89.69</td>
<td>13.54</td>
<td>82.81</td>
<td>16.43</td>
<td>91.56</td>
<td>9.55</td>
<td>85.31</td>
<td>12.94</td>
<td>87.34</td>
<td>9.62</td>
</tr>
<tr>
<td>Overall</td>
<td>90.16</td>
<td>10.86</td>
<td>85.63</td>
<td>11.77</td>
<td>87.81</td>
<td>11.14</td>
<td>88.44</td>
<td>10.26</td>
<td>88.01</td>
<td>8.77</td>
</tr>
</tbody>
</table>
However, the effect of position was significant, $F(3, 117) = 60.649$, $p < 0.001$, $\text{MS}_{\text{position}} = 983214.246$, as was the interaction effect, $F(3, 117) = 7.063$, $p < 0.001$, $\text{MS}_{\text{interaction}} = 82793.704$ (see Table 6 and Fig. 3). Simple effect tests showed that the effect of position was significant with both blue links, $F(3, 117) = 44.665$, $p < 0.005$, $\text{MS}_{\text{position}} = 596807.082$ and red links, $F(3, 117) = 32.199$, $p < 0.005$, $\text{MS}_{\text{position}} = 469200.869$. Further simple effect tests showed that with top presentations, blue links were faster than red links ($p < 0.001$). Simple comparison tests showed that with blue links the top and the bottom position were faster than both the right position ($p < 0.001$ and 0.02, respectively) and the left position ($p < 0.001$ and $p = 0.001$, respectively); the top position was also significantly faster than the bottom position ($p < 0.001$). Subsequent simple comparison tests showed that with red links the right position was faster than the left ($p < 0.05$) and both the top and bottom positions were faster than the right and left positions (all four $p < 0.001$).

### Table 4
Descriptive statistics for reaction time for hits (visual search task)

<table>
<thead>
<tr>
<th>Link colour</th>
<th>Right</th>
<th>Left</th>
<th>Top</th>
<th>Bottom</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
</tr>
<tr>
<td>Blue</td>
<td>1405</td>
<td>220</td>
<td>1348</td>
<td>239</td>
<td>1244</td>
</tr>
<tr>
<td>Red</td>
<td>1386</td>
<td>270</td>
<td>1421</td>
<td>220</td>
<td>1211</td>
</tr>
<tr>
<td>Overall</td>
<td>1396</td>
<td>231</td>
<td>1384</td>
<td>209</td>
<td>1228</td>
</tr>
</tbody>
</table>

*Note: milliseconds.*

### Table 5
Descriptive statistics for percentage of hits (visual search task)

<table>
<thead>
<tr>
<th>Link colour</th>
<th>Right</th>
<th>Left</th>
<th>Top</th>
<th>Bottom</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
</tr>
<tr>
<td>Blue</td>
<td>95.94</td>
<td>6.57</td>
<td>97.50</td>
<td>5.80</td>
<td>98.13</td>
</tr>
<tr>
<td>Red</td>
<td>96.88</td>
<td>6.79</td>
<td>97.81</td>
<td>4.81</td>
<td>95.63</td>
</tr>
<tr>
<td>Overall</td>
<td>96.41</td>
<td>5.08</td>
<td>97.66</td>
<td>3.66</td>
<td>96.88</td>
</tr>
</tbody>
</table>

### Table 6
Descriptive statistics for reaction time for correct rejections (visual search task)

<table>
<thead>
<tr>
<th>Link colour</th>
<th>Right</th>
<th>Left</th>
<th>Top</th>
<th>Bottom</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
<td>S.D.</td>
<td>$M$</td>
</tr>
<tr>
<td>Blue</td>
<td>1582</td>
<td>288</td>
<td>1612</td>
<td>274</td>
<td>1340</td>
</tr>
<tr>
<td>Red</td>
<td>1573</td>
<td>291</td>
<td>1650</td>
<td>263</td>
<td>1431</td>
</tr>
<tr>
<td>Overall</td>
<td>1577</td>
<td>280</td>
<td>1631</td>
<td>259</td>
<td>1385</td>
</tr>
</tbody>
</table>

*Note: milliseconds.*
3.2. Results for the interactive search task

3.2.1. Incorrect answers

Percentage of errors in answers for the interactive search task was calculated and results appeared to indicate a ceiling effect (see Table 7 and Fig. 4). Position had no effect on performance, $F(3, 117) = 1.31$, $p > 0.05$. However, the effect of colour, $F(3, 117) = 4.738$, $p < 0.05$, $\text{MS}_{\text{colour}} = 5.157$ and the interaction effect were significant, $F(3, 117) = 3.719$, $p < 0.05$, $\text{MS}_{\text{interaction}} = 4.995$. Simple effect tests showed that the effect of position was not significant with blue links, $F(3, 117) = 2.686$, $p > 0.05$, $\text{MS}_{\text{link}} = 4.720$ and not significant either with red links, $F = 1$. Further simple effect tests showed that the effect of colour was significant for links presented at the right of the screen ($p < 0.05$), with blue links producing fewer errors than red links.

3.2.2. Time-on-task

Time-on-task for correct answers was calculated. The effects of position, $F(3, 117) = 8.410$, $p < 0.001$, $\text{MS}_{\text{position}} = 61.015$, colour, $F(1, 117) = 14.526$, $p < 0.001$, $\text{MS}_{\text{colour}} = 89.866$ and the interaction effect, $F(3, 117) = 9.154$, $p < 0.001$, $\text{MS}_{\text{interaction}} = 65.341$, were all significant (see Table 8 and Fig. 5). Simple effect tests showed that the effect of position was significant with blue links, $F(3, 117) = 6.691$, $p < 0.001$, $\text{MS}_{\text{position}} = 33.467$ and with red links, $F(3, 117) = 9.891$, $p < 0.001$.

### Table 7

<table>
<thead>
<tr>
<th>Link colour</th>
<th>Right</th>
<th>Left</th>
<th>Top</th>
<th>Bottom</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>0.23</td>
<td>0.83</td>
<td>0.23</td>
<td>1.09</td>
<td>0.47</td>
</tr>
<tr>
<td>Red</td>
<td>1.02</td>
<td>1.79</td>
<td>0.78</td>
<td>1.37</td>
<td>0.47</td>
</tr>
<tr>
<td>Overall</td>
<td>0.63</td>
<td>1.11</td>
<td>0.51</td>
<td>0.89</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Fig. 3. Interaction between position and link colour for reaction time for correct rejections (visual search task).
Further simple effect tests showed that the effect of colour was significant for links presented at the left and the top of the screen (both $p < 0.005$), with blue links producing faster responses than red links. Simple comparison tests with Sidak correction showed that, with blue links, the left position was faster than both the top ($p < 0.05$) and bottom positions ($p < 0.005$) and the right position was faster than the bottom position ($p < 0.05$). Further simple comparison tests showed

\[ MS_{\text{position}} = 92.889. \]
that, with red links, the top position was slower than both the right \((p = 0.001)\) and bottom positions \((p < 0.001)\).

3.3. Results for the visual search task and the interactive search task

Correlations between corresponding performance measures in the visual search task and the interactive search task were calculated. It was found that none of the eight correlations between percentages of hits (visual search task) and the eight percentages of errors (interactive search task) were significantly correlated, nor were the eight reaction times for hits (visual search task) and the eight times-on-task for correct answers (interactive search task).

The most preferred location for the navigation frame was the left of the screen (50%); second and third most preferred were the top (20%) and the right positions (17.5%); the least favoured location was bottom (12.5%). A chi square test of independence showed these differences to be significant, \(\chi^2(3, n = 40) = 13.800, p < 0.005\). Preferences for link colour were 57.5% for blue and 42.5% for red, a non-significant difference, \(\chi^2(1, n = 40) = 0.343, p > 0.05\).

Finally, comparisons were made between those who preferred the left position most and those who preferred other positions in their performance on both the visual search task and the interactive search task in trials where the navigation menu was presented on the left. Out of 12 tests, \(^2\) one was significant (correct rejections with red links, \(t(27.177) = 2.147, 0.04 < p < 0.05\), without correction for conducting multiple comparisons).

4. Discussion

4.1. Visual search task

The experimental hypothesis that the means for the four levels of presentation position would differ was supported, for measures of speed (reaction times for hits and correct rejections), but not for measures of accuracy (percentages of hits and correct rejections). For reaction times for correct rejections the interaction effect between presentation position and link colour was significant.

For both the blue and red link colours the reaction time for hits was significantly faster for the top and bottom presentation positions compared with the right and left. This was also found to be the case for the reaction time for correct rejections, but top was also found to be significantly faster than bottom for a blue link colour and right significantly faster than left for a red link colour.

The results for measures of speed do not conform to those found by van Schaik and Ling (2001a, b). They used search tasks with the same type of screens and found that (a) top and left positions were fastest with hits and (b) no difference existed with

\(^2\) \((4\) performance measures for the visual search task \(+ 2\) performance measures for the interactive search task) \(\times 2\) link colours.
correct rejections. However, differences in findings may be due to the fact that the current study used unrelated link words that were approximately matched for word frequency whereas van Schaik and Ling used sets of related link words that represented nine common concepts investigated by Barsalou (1983) without matching for word frequency. Arguably, van Schaik and Ling’s link words are more realistic because in web pages lists of links are usually related rather than unrelated.

These findings also contradict guidelines which state that menu items should be placed vertically (Tullis, 1997) and research that found search time was shorter for vertical menus (Wolf, 1986; Backs et al., 1987; Nygren, 1996). Furthermore, they contradict the view that menu items should be placed on the left (Campbell and Maglio, 1999; Nielsen, 1999a) or right (Nielsen, 1999a), as well as the idea that important information should not be placed at the bottom of the screen as people ignore it (Spool, 1999).

The research of Scott and Findlay (1990) is supported by these findings as the search time for horizontal menus was found to be significantly faster than that of vertical ones. These findings also support the idea that navigation should be placed at the top and bottom of screens (Borges et al., 1998).

The top presentation position being the fastest does support, to some extent, the view of Galitz (1997), who states that people look at the upper left of the display, then quickly move through the display in a clockwise direction. The finding that the bottom position is significantly faster than the right, however, suggests that this idea may be not be completely correct.

The fact that presentation position did not have an effect on the number of correct responses in the visual search task does not support the findings that horizontal menus produce significantly more errors than vertical ones (Yamamoto and Kuto, 1992).

For hypertext link colour, the experimental hypothesis was not supported for any of the performance measures. This finding is contradictory to the guidelines and findings that blue is a poor colour for text (Murch, 1984; Galitz, 1997; Marcus, 1997; Nielsen, 1999a) and that red should be used to draw a user’s attention (Galitz, 1997). It also contradicts the opposite view that on a white background blue is one of the most effective colours for text (Lalomi and Happ, 1987; White, 1990; Fowler and Stanwick, 1995).

Pastoor (1990) and Pace (1984) are supported by this result as they found no significant difference for performance of different character background combinations. In this case though this may be due to and perhaps add weight to, the idea that “when fast responses are needed, use highly saturated red or blue prompts” (Durrett and Trezona, 1982). The conjecture may therefore be proposed that both blue and red text result in the same level of performance, one that is better than that of other colours when fast responses are required.

4.2. Interactive search task

The experimental hypothesis that the means for the four levels of presentation position would differ, was supported for time-on-task, but not for error rate.
For the blue link colour, time-on-task for the left presentation position was significantly faster than that of the bottom and top positions. Time-on-task for the right position was also found to be significantly faster than the bottom position.

For the red link colour the results were different. In this case the right, left and bottom positions times-on-task were found not to significantly differ, but the right and bottom positions were found to be significantly quicker than that of the top position.

These findings, for both red and blue links, do therefore partly support guidelines that state that menu items should be placed vertically (Tullis, 1997) and research that found search time was shorter for vertical menus (Wolf, 1986; Backs et al., 1987; Nygren, 1996). They therefore partly contradict the findings of Scott and Findlay (1990) that horizontal menus are searched more quickly than vertical ones.

The position of the menu frame on the screen does seem to be the major factor in determining time-on-task. For a blue link colour these findings support the idea that menu items should be placed on the left of the screen (Campbell and Maglio, 1999; Nielsen, 1999a) and give slight support for them to be placed on the right (Nielsen, 1999a). The view of Borges et al. (1998) that navigation should be placed on the bottom and top of the screen is contradicted by these findings and therefore the idea that information should not be placed at the bottom of the screen, as people ignore it (Spool, 1999) is supported.

For red links however, Spool (1999) view is not supported as no significant difference was found between time-on-task for the bottom, left and right positions, with the significant difference occurring between top and the right and bottom positions. This therefore means that the views that the best presentation position should be the left (Campbell and Maglio, 1999; Nielsen, 1999a), the right (Nielsen, 1999a), or the bottom (Borges et al., 1998) are neither completely contradicted or supported as time-on-task for these positions did not differ. The view that menu items should be placed at the top (Borges et al., 1998) though is definitely not supported.

No evidence was found for an overall effect of position on error rate. So the findings that vertical menus produce fewer errors (Yamamoto and Kuto, 1992) is not supported.

For hypertext link colour, the experimental hypothesis was supported for both performance measures. Blue links resulted in fewer errors with right presentations and faster correct responses with left and top presentations. This means that the guidelines and findings that blue is a poor colour for text (Murch, 1984; Galitz, 1997; Marcus, 1997; Nielsen, 1999a) and that red should be used to draw a user's attention (Galitz, 1997) are not supported. This result is also contradictory to the view that there is no difference in performance for different character background combinations (Durrett and Trezona, 1982; Pace, 1984; Pastoor, 1990). It does therefore support the view that on a white background blue is an effective colour for text (Lalomia and Happ, 1987; White, 1990; Fowler and Stanwick, 1995).

This finding also challenges the idea that blue links may distract users from other tasks (due to them being easily seen in the periphery; Murch, 1984) and so cause a
user’s attention to be divided (Marks and Dulaney, 1998), resulting in attention to both items suffering (Neisser and Becklen, 1975).

During the analysis of the videotape of the interactive search task it was observed that a majority of participants might have used the mouse cursor either as an aid for directing their attention (aiding peripheral vision), or an aid for reading (aiding foveal vision), as it was moved towards or onto information that they seemed to be processing.

Without further investigation it is not possible to determine whether the use of the cursor as a visual aid (either peripheral or foveal) affected time-on-task. However, this perhaps suggests that an additional method of making targets (links) that are close to the cursor pop up and particularly distinct may aid web page usability, especially for people with visual disabilities. This could be done for example by changing link colour on proximity of the cursor. Indeed, this type of highlighting is supported by web browsers that include an optional setting with the effect of changing the link colour when the cursor “hovers” over a link. Cascading style sheets provide more flexibility in supporting this type of highlighting by supporting, for example, control over text size and background colour of link words, in addition to change in colour. In order to investigate the use of the cursor as a visual aid further, eye movements could be recorded using eye-tracking equipment.

4.3. Comparison of the two tasks

As can be seen from above, the results of the two tasks do seem to differ. Firstly, there was no effect of link colour in the visual search task, whereas blue was found to be significantly quicker than red for the interactive search task. Secondly, the best performing presentation positions, in terms of time, for the visual search task were the top and bottom, whereas for the interactive search task, especially for the blue link colour, they were the left and right positions.

The non-significant results from the correlations for both measures of speed and measures of accuracy also support the idea that the two tasks may be measuring fundamentally distinct types of task performance.

The differing results of the two tasks can be explained in two ways: (1) the visual search task may lack external validity or (2) the interactive search task may be systematically biased in some way.

The visual search task may lack external validity due to it having no real utility (Comber and Maltby, 1997). Even though the web screens that participants had to search were in fact screens dumps of HTML web pages, the task that participants had to carry out with the screens (either pressing “a” or “p”) was not closely related to what a user would do when searching for information on the Web as in the interactive search task. This could mean that automatic attention responses that may have formed about web sites for experienced Internet users, as is the case for the majority of participants, would not be applied. For example the automatic attention response that blue writing means a hypertext link (Marks and Dulaney, 1998; Nielsen, 1999a) and that links are placed on the left (Nielsen, 1999a) would not be...
applied, as seems to be the case and therefore would not aid performance. It would therefore be expected that other strategies would be used to aid performance. This may in fact explain why horizontal menus (top and bottom) had a significantly quicker reaction time than vertical menus (right and left) for both hits and correct rejections for the visual search task. The reason for this result may be that horizontal arrays are conforming to left-to-right readings habits that are prevalent in Western cultures (Scott, 1993), but not to vertically presented web page menus.

For the interactive search task, as actual web pages were used, one would expect participants to use these automatic attention responses. This seems to be the case, firstly as time-on-task for questions with blue links (default link colour) was found to be significantly quicker than those of red links and secondly because for blue links the left (default presentation position) was significantly quicker than both the top and bottom presentation positions. Future research should investigate if these automatic attention responses are actually formed by Internet users as they become more experienced.

Support for the idea that automatic attention responses about web pages may have been applied by participants in the interactive search task comes from the fact that a significant difference between frequencies was found for preferred presentation position, with half of the participants stating they preferred the left presentation position. However, preference for the left position was not predictive of performance level on pages with the navigation menu presented on the left; this reinforces Nielsen and Levy’s (1994) finding of notable exceptions to their general finding of a positive correlation between performance and preference measures in human–computer interaction research studies.

More participants stated that blue was their preferred link colour but the difference in frequencies was not found to be significant. The non-significant difference for link colour may be due to the fact that participants may have preferred red for visual search task but blue for the interactive search task, as was stated by a number of participants, who were therefore unsure how to respond. This actually adds support to the idea that participants may have used different automatic attention responses in the two tasks. This would have to be investigated further though, for example, by asking participants their preference after each task and monitoring their actual eye movements with an eye tracker.

This idea that the visual search task lacks external validity, due to it not causing relevant web page-related automatic attention responses to be applied, supports the view that real users should be observed carrying out interactive search tasks (Bevan and McLeod, 1994; Spool et al., 1997; Nielsen, 1999b) and the view of Carroll (1989) that psychological experiments are not always representative of real problems.

There are, however, limitations of the interactive search task that may have affected results. Firstly, differing levels of cognitive load could have affected the results across the questions. Support for this comes from Graf and Krueger (1989) who found that performance on a task similar to that of this experiment (which consisted of both reading and searching components) was affected by the difficulty of the question. Some questions in the interactive search task were longer than others
and some required participants to remember more if they were to find the answer. The content of questions could also have an effect. However, effects of difficulty and content may have cancelled out between different combinations of presentation and colour. In order to minimize this effect future research could use a mixed design or independent measures design, with each question being presented for each position and link colour at random to different participants.

Secondly, the number of questions that participants answered for a mean to be calculated for each variable (four) may have been too low. Future research may therefore increase the number of questions that participants were given.

Thirdly, Nielsen’s (1999a) claim that the right-hand side of the screen would be the best for links, due to them being nearer the scroll bar, may not have been supported due to the complex path through which participants had to move the cursor. They had to move the cursor a large distance to acquire targets when the menu frame was on the right and the question required participants to go through two levels of links and scroll the screen. This was because after clicking on the first link participants had to move the cursor right across the screen to click on the second level of links, presented on the left of the screen and then back again to acquire the scroll bar. Because of this, the overall distance travelled for these questions was probably equal or greater than the distance travelled for the other questions.

The interactive search task appears to mimic real-world tasks more closely and have more real utility than the visual search task, and so perhaps measured what it was setting out to measure more thoroughly. It is still however experimental in nature, so rather than suggesting that controlled psychological experiments do not have a place in human–computer interaction research (Winograd and Flores, 1986; Whiteside and Wixon, 1987; Carroll, 1989), this research suggests that more thought has to be put into the designing of measures for psychological experiments so that they have real utility and cannot be classed as toy problems.

4.4. Methodological issues

Related to the current study, Diaper and Waelend (2000, p. 166) set out to address the “failure” of “the application of theoretically oriented and applied psychological laboratory research to the real world of web page usage”. The authors investigated the effect of animation on task performance and perceived complexity of web pages and contrasted their experiments with those conducted by Zhang (1999). In essence, Diaper and Waelend concluded from their research that animation does not affect users’ task performance, whereas Zhang found that this adversely affected performance. Zhang used a visual search task with arrays of meaningless letter strings and blocks of random words. Diaper and Waelend, however, employed what was essentially a visual search task using a modification of an existing web page consisting of text, but the task did not involve interaction with the web page. In terms of ecological validity, because it involved realistic web pages Diaper and Waelend’s task is arguably more realistic than Zhang’s and the visual search task used in the current study. However, the interactive search task used in the current study is more valid because it involved interaction, whereas Diaper and Waelend’s
An advantage of the current study is that participants carried out both the visual search task and the interactive search task, making results between tasks directly comparable. Diaper and Waelend’s comparison of their results with those of Zhang lacks validity in the sense that (1) different samples were used that were probably drawn from different populations and (2) most likely the experimental setting and equipment was different.

Considerable care is required when drawing conclusions from experimental research as a basis for guidelines. This is not only true in cases where the experimental task is a more basic task such as visual search, but also where a task is used that is apparently more realistic. A number of flaws found in the Diaper and Waelend (2000) study highlight the need for care in research studies more generally. A lack of external validity can occur when web pages that are used are not from the domain (for example advertising) to which the research seeks to generalize. A failure to analyse and interpret a statistical interaction effect can seriously degrade the validity of conclusions from a research study. A failure to employ counterbalancing (where this is crucial to the argument of the research) can undermine any conclusions that are drawn from research results. Before measures of perceived quality of display can be recommended they need to be validated by establishing correlations with measures of task performance or other relevant measures (Ahuja and Webster, 2001). An assessment should be given of how realistic the experimental task is compared to other possible tasks (for example a visual search task compared to an interactive search task). It is particularly important to avoid sending incorrect and arguably dangerous messages to designers who require advice on designing their web pages; more specifically, it would appear to be dangerous to tell designers that they can use their own judgement when making decisions about choices of web design parameters.

4.5. Guidelines

Even if the results from the interactive search task from the current study are taken as being reliable it is still not really possible to give definitive guidelines relating to hypertext link colour and presentation position of navigation menus without further investigation.

It is valid to recommend, however, leaving the default link colour as blue, when using a white background, rather than changing it to red. This does not mean that blue would be best for a different background colour and that there is not a better other colour than red. In order to test this, future research would have to examine a large number of character background colour combinations.

If blue links are retained for a white background then it can be stated that the optimum position for a navigation menu is the left, with the right being seen as acceptable. On the other hand if red links are employed on a white background then the only guideline that can be given is that the navigation menu should not be placed at the top of the screen.

The results from the visual search task however show that search time is quicker for horizontal menus (top and bottom positions) and that in this case participants
may not have used specific web page-related automatic attention responses. From this it would be predicted that for novice participants the result would be the same for the visual search task as for the interactive search task, as they would not have specific web page-related automatic attention responses. If this is the case then for novice users, at least from Western cultures, the placing of the navigation menu at the top or bottom of web pages would be advantageous as it would improve initial performance, as users would not have to spend time learning new automatic attention responses. Due to the steadily increasing number of Web users, if a change to design conventions is to be made it would be better to be sooner rather than later, as a greater number of experienced users would have to learn new automatic attention responses the later a change was made (Nielsen, 1999c). Future research therefore should investigate how automatic attention responses form for novice users and also the ability of experienced users to alter pre-formed automatic attention responses.

5. Future work

Task performance is an important type of outcome measure of human–computer interaction in web sites, but not the only relevant one. Validation of behavioural measures and psychometric measures with task performance (Ahuja and Webster, 2001) is important. For example, Ahuja and Webster, in their landmark study, found that perceived disorientation was predictive of task performance in an interactive search task, but users’ actions were not. In addition to correlations between behavioural (actions carried out with a web site) and psychometric measures, think-aloud protocols (Nielsen, 1993) can be analysed to provide explanations for different levels of task performance. This can be useful in particular to find causes of and solutions to usability problems.

Further explanation for differences in outcomes between the visual search tasks and interactive search tasks can be derived from a detailed analysis of the structure of the two types of task. In terms of information processing, the visual search task—using the choice reaction paradigm—can be thought of as consisting of four consecutive stages (van Schaik and Ling, 2001b): (1) a global search of the whole screen to locate the navigation frame, (2) a local search of the navigation frame to locate the target word, (3) a decision to make a motor response indicating either the presence or absence of the target and (4) making the motor response to show that a target has been detected if Stage 2 was successful or that no target has been detected if Stage 2 was not successful.

The interactive search task can be thought of as consisting of four consecutive stages: (1) a global search of the whole screen to locate the navigation frame on the first page, (2) a sequence of local search actions in the first and following pages, involving (a) a local search of the navigation frame to locate a matching word, (b) move to the target and (c) clicking the target, (3) a search for the target and the answer in the content area of the page that contains the answer, (4) move to the answer box and (5) making a typing response. Following Gray and Boehm-Davis’s
(2000) approach, using cognitive-perceptual-motor modelling times taken by the different stages can be calculated (including the use of Fitts’ law for movement time) and estimates can be made of the effect of presentation position and links colour on the stages. Unfortunately, the method of data collection used in the interactive search task in the current study—recording only the total time-on-task but not times for the separate stages—did not allow these calculations and estimates to be made. However, our on-going programme of research in this area now includes automatic time-stamped logging of all visited web pages (van Schaik and Ling, in press) and we recommend the cognitive-perceptual-motor paradigm as a useful tool to be considered in further research.

6. Conclusion

In conclusion, the current study employed a promising approach for experimentally studying the effects of multiple design parameters on task performance, using the paradigm of interactive search. Experimental results from an interactive search task differed from those produced by a visual search task. The interactive search paradigm is arguably more ecologically valid than non-interactive search in web pages (Diaper and Waelend, 2000), visual search with meaningful stimuli (van Schaik and Ling, 2001a,b) and visual search with meaningless or random stimuli (Zhang, 1999). This study found some support for automatic attention responses and guidelines that are consistent with the notion of automatic attention responses. The need for on-going research to further develop knowledge of human–computer interaction on the Web and Web design guidelines has been confirmed by Nielsen (2002): “There are more papers on unworkable, esoteric 3-D browsers than on how hundreds of millions of people use the biggest real-time collaborative system ever built” [the World Wide Web].

References


