CARDIFF UNIVERSITY

An Empirical Investigation into Strategies for Guiding Interactive Search

Duncan P. Brumby

8th December 2005

School of Psychology

Cardiff University

Cardiff, Wales, UK CF10 3AT

A thesis submitted for the degree of

Doctor of Philosophy

Thesis advisor: Andrew Howes

Viva voce examination committee:

Frank Ritter (external examiner)

Ulrike Hahn (internal examiner)

Rob Honey (chair)

CONTENTS

FIGURE CAPTIONS

TABLE CAPTIONS

PUBLICATIONS

Experiment 4 from Chapter 3 has previously been presented:

Brumby, D. P. & Howes, A. (2003). Adaptive decision making in menu search: The role of interdependence and past experience on link selection. In P. Gray, H. Johnson, & E. O'Neill (Eds.), *Proceedings of HCI 2003: Designing for Society Volume 2* (pp. 133-134), London, UK: Springer-Verlag.

Brumby, D. P. & Howes, A. (2003). Interdependence and past experience in menu choice assessment. Poster session presented at the *25th Annual Conference of the Cognitive Science Society,* Boston, MA.

The ACT-R model described in Chapter 4 and brief report of Experiment 1 from Chapter 3 has previously been published:

Brumby, D. P. & Howes, A. (2004). Good enough but I'll just check: Web-page search as attentional refocusing. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), Proceedings of the *6th International Conference on Cognitive Modelling* (pp. 46-50), Mahwah, NJ: Lawrence Erlbaum.

Brumby, D. P. (2004). A model of single-page web search: The effect of interdependence on link assessment. In M. Lovett, C. Schunn, C. Lebiere, & P. Munro (Eds.), Proceedings of the *6th International Conference on Cognitive Modelling* (pp. 402-403), Mahwah, NJ: Lawrence Erlbaum.

Manuscript currently under review:

Brumby, D. P. & Howes, A. (submitted). Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *Human-Computer Interaction.*

ACKNOWLEDGMENTS

Dedicated to the loving memory of Muriel Brumby and Evelyn Lindsey Newcombe who will be warmly remembered and deeply missed.

It has been an exciting couple of years, although somewhat tumultuous of late. I would like to thank everyone in my family for their continuous support. Megan Brumby is a star.

I am grateful to have been supported throughout my graduate studies by an EPSRC studentship given to the School of Psychology at Cardiff University.

Many people at the School of Psychology at Cardiff University have made the past few years academically stimulating and wholly enjoyable. I am especially grateful to Andrew Howes, my thesis advisor, for providing continuous support and guidance throughout the development of the work presented in this thesis. I would also like to thank Richard Young and Anna Cox at UCL for their collaboration over the course of conducting this research and in particular for kindly providing me with the source code for their model.

The work presented in this thesis has substantially befitted from extensive discussion with others during its numerous outings. In particular, I am grateful to: Alonso Vera and Irene Tollinger for inviting, and then stepping through the odious paperwork, so that I could visit NASA Ames Research Center; Anthony Hornof for inviting me to the University of Oregon; Richard Young for making those interesting in ACT-R in the UK aware of each others' existence, and organizing a small group meeting that proved to be particularly helpful to the development of the model presented in this thesis; All those that took part in the Doctoral Consortium workshops at HCI 2003 in Bath, and ICCM 2004 in Pittsburgh, particularly, the organizers Ann Blandford and Lachlan MacKinnon (HCI 2003) and Wayne Gray and Dave Touretzky (ICCM 2004); Where would things be without the annual Gregynog experience?

Finally, I feel very lucky, and indeed privileged, to have had two truly remarkable summers away from Cardiff working as an intern; first at Palo Alto Research Center (PARC) with Peter Pirolli and Stuart Card in 2004, and at Microsoft Research with Ed Cutrell, Raman Sarin, Susan Dumais and Ken Hinckley in 2005.

ABSTRACT

One activity people engage in when using the web is estimating the likelihood that labelled links will lead to their goal. However, people must also decide whether to select one of the assessed items immediately or make further assessments. There are a number of theoretical accounts of this behaviour. The accounts differ as to whether, for example, they assume that people consider all of the items on a page prior to making a selection, or tend to make a selection immediately following an assessment of a highly relevant item. A series of experiments were conducted to discriminate between these accounts. The empirical studies demonstrated that people are in fact more strategic and sensitive to context than previous models suggest. People sometimes choose an option that appears good enough, but sometimes choose to continue checking. The decision to select an item was found to be sensitive to the relevance of labels in the immediate and distal choice set and also the number of options in the immediately available choice set. The data were used to motivate computational models of interactive search. The development of explicit, formal cognitive models of how people search web sites holds the potential to provide clear and unambiguous information to support future web design and improve usability.

CHAPTER 1: INTRODUCTION

How do people search a newly encountered web page for a link that is relevant to the achievement of their search goal? It is known that estimates of label relevance play a substantial role (Card, Pirolli, Van der Wage, Morrision, Reeder, Schraedley, & Boshart, 2001; Chi, Pirolli, Chen, & Pitkow, 2001; Chi, Pirolli, & Pitkow, 2000; Pirolli & Fu, 2003; Miller & Remington, 2004). Others have made the case that the design of a web site might be improved by considering the relevance of items on its constituent pages (Blackmon, Polson, Kitajima, & Lewis, 2002; Blackmon, Kitajima, & Polson, 2003; Chi, Rosien, Suppattanasiri, Williams, Royer, Chow, Robles, Dalal, Chen, & Cousins, 2003; Kaur & Hornof, 2005). One neglected issue, however, is that during search a user must decide which items to assess and how to assess them. That is, estimates of relevance must be embedded within a strategy for controlling search.

The aim of this thesis was to investigate strategies for guiding search of a novel web page for information that is relevant to the achievement of a particular search goal. Following Payne, Howes, and Reader (2001, pp. 340) I will refer to this task as *interactive search*. Generally, interactive search tasks require the user to successful discrimination between items that actually lead to the goal (*target items*) from those that do not lead to the goal (*distractor items*). A particular focus of the work presented in this thesis was on how people decide between selecting one of the assessed items and continuing to make further assessments.

The thesis is organized as follows. Chapter 2 reviews previous empirical studies of interactive search that have focused on regularities in how people search web pages (e.g., Card et al., 2001) and also database menu pages (e.g., MacGregor, Lee & Lam, 1986; Pierce, Parkinson & Sisson, 1992). While there is limited empirical evidence concerning how search behaviour might be controlled, there is a substantial theoretical literature in which a number of cognitive models of interactive search have been proposed (e.g., Cox & Young, 2004; Howes, 1994; Howes, Payne, & Richardson, 2002; Lee & MacGregor, 1985; MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Rieman, Young, & Howes, 1996; Young, 1998). Each model makes quite different theoretical assumptions about how people choose between assessment and selection. To foreshadow the review, there are three possible strategies people could adopt during interactive search. People can either: 1. Consider all of the items on a page prior to making a selection (i.e., take the best, without a selection threshold); or 2. Immediately select an item, if some estimate of an item's relevance is greater than an arbitrary selection threshold (i.e., a simple selection threshold); or 3. The decision to whether to select an item might be sensitive to the entire set of assessments made (i.e., a context determined dynamic selection threshold).

Chapter 3 presents a series of experiments that were designed to discriminate between the three main views of the strategies that people adopt during interactive search. The experiments systematically manipulated the relevance of the target and distractor items, and the location of the target item within the set. An additional experiment considered the potential role of distal influences on behaviour during interactive search. The experiments provided no evidence to support the hypothesis that people persistently assess every item prior to deciding which to select, nor for the idea that people assess items until the most recent exceeds a threshold. Instead it was found that the decision to select an item was sensitive to the estimated relevance of all of the items so far assessed, and not just to the most recent item.

The aim of Chapter 4 was to explore how and also whether cognitive models of interactive search can account for the main experimental results reported in the previous chapter. In particular, context sensitive models of interactive search rely on a *normalization* assumption (Young, 1998) in order for the subjective value of selecting an item to be sensitive to the context provided by the previously visited item in the choice set. A cognitive model of interactive search was developed that was inspired by Young's normalization assumption, but which was also sensitive to the psychological constraints encoded in the ACT-R theory of the human cognitive architecture (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004). The model implemented interactive search as attentional focusing. In the model the probability of successfully retrieving a declarative chunk associating a labelled link with the goal statement was partly dependent on the number of other labels in the choice set which were also relevant to the goal.

One area in which models of interactive search have been applied is to consider the optimal structure of a web site (i.e., whether to prefer broad and shallow or deep and narrow choice sets in a web site design). The experiments in Chapter 5 attempted to tease apart predictions derived from previous models of interactive search (e.g., ACT-R model, Chapter 4; Cox & Young, 2004; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003) of the affect of varying the number of labelled links on a single web page for assessment and selection. This was particularly pertinent because Miller and Remington's (2004) model, which claims to predict the average time to navigate a web site, assumes that search of each page was based on a threshold strategy. A threshold account failed to capture performance on the single page search tasks studied in Chapter 3.

In summary, the work presented in this thesis aims to investigate strategies for guiding interactive search. It might be the case that when navigating the web people simply select the link on a page with the highest scent, as claimed by Pirolli and Card (1999), or they might simply assess items until the most recent exceeds a threshold, as claimed by Miller and Remington (2004). The results of the empirical studies suggest that people are in fact more strategic and sensitive to context than previous models suggest. Decisions are continually made about whether to select one of the assessed items immediately or whether to make further assessments and each decision is sensitive to the estimated relevance of all of the items so far assessed, and not just to the most recent item. The theoretical work of modelling these key findings allows for explicit theories of interactive search (e.g., Cox & Young, 2004; Howes, Payne, & Richardson, 2002; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Rieman, Young & Howes, 1996; Young, 1998) to be formulated and evaluated.

CHAPTER 2: LITERATURE REVIEW

2.1. How do People Search Web Pages?

While the focus of this thesis is very much on the strategies people deploy to choose between assessment and selection of labelled web links, there is a large literature on the more general topic of how people search web pages and sites. This literature provides an invaluable context against which the role of relevance should be set. For example, people use the web to fulfil a variety of everyday needs, such as navigating links and by using a search engine (Morrison, Pirolli, & Card, 2001; Byrne, John, Wehrle, & Crow, 1999; Cockburn & McKenzie, 2001; Sellen, Murphy, & Shaw, 2002). People often fail to go directly to a site or page that satisfies their goal and they use the backup button reasonably frequently (Catledge & Pitkow, 1995). Considering the design of a web site, e.g., its depth/breadth can affect how quickly people satisfy goals (Katz & Byrne, 2003; Larson & Czerwinski, 1998; Norman, 1991; Parush & Yuviler-Gavish, 2003; Miller & Remington, 2004; Snowberry, Parkinson, & Sisson, 1983). The spatial layout of the page has consequences for ease of navigation (McCarthy, Sasse, & Riegelsberger, 2003) as can the colour of the hyperlinks (Halverson & Hornof, 2004; Pearson & Schaik, 2003).

Search engines provide a powerful tool for finding goal relevant information on the web. A number of studies have analysed query terms and link selection data from search engine user logs (Jansen, & Pooch, 2000; Lau & Horvitz, 1999; Mat-Hassan & Levene, 2005; Silverstein, Henzinger, Marais, & Moricz, 1998; Spink, Bateman, & Jansen, 1999). People tend to submit queries to search engines, which contain only a few words (approx. 3 words) (Spink, Bateman, & Jansen, 1999). Query terms are frequently reformulated during the course of a single search session, however, as people narrow-in or expand on their search topic (Lau & Horvitz, 1999; Spink, Bateman, & Jansen, 1999). When exploring the results list, Jansen and Pooch (2000) and Silverstein et al. (1999) found that people select very few links over the duration of the search session, and rarely even go beyond the first page of the results list. Mat-Hassan and Levene (2005) found that as the duration of a single search session increases, however, people are more likely to explore links further down in the search engine ranking (i.e., results beyond the first page are explored).

The relevance of links to a user's particular information goal, is one issue that has received considerable empirical and theoretical attention within the literature (Blackmon, Polson, Kitajima, & Lewis, 2002; Blackmon, Kitajima, & Polson, 2003; Card et al., 2001; Chi et al., 2003, 2001, 2000; Church & Keane, 2004; Katz & Byrne, 2003; Kaur, & Hornof, 2005; Miller & Remington, 2004; Pierce, Parkinson, & Sisson, 1992; Pirolli & Fu, 2003). Unsurprisingly, users tend to select items (i.e., labelled links) from web pages that are relevant to their goal (Card et al., 2001; Katz & Byrne, 2003; Miller & Remington, 2004). Card et al. (2001) observed participants while engaged in goal-directed search of the web. In the study, participants searched for information relevant to the achievement of ecologically valid goals. A user trace was constructed, based on eye-tracking data, application-level logs, and verbal protocols. They found that participants were likely to select items from a web page that were of greater semantic relevance to their information goal. This scent following strategy was particularly evident from verbal protocols. It was also found that when a participant traversed a number of pages within a web site, they tended to switch to a different web site (i.e., leave a site), when the relevance of the items in the site decreased below some typically experienced value.

Label relevance also influences navigation choices within a web site. In a study by Katz and Byrne (2003), participants searched toy-web sites to locate information that was relevant to a given goal. In the experiment, participants could opt to navigate the site by either using an in-built search feature (i.e., similar to a Google search) or by browsing a menu structure. Katz ad Byrne found that the decision between using the search feature or browsing a menu structure was influenced by the number of items on a particular web page and the semantic relevance of the items; when a page contained more items of lower relevance to the search goal participants were more likely to use the search feature.

Although users tend to select items that are relevant to their goal, one important issue to consider is the ease with which they can successfully discriminate between items that actually lead to the goal (*target items*) from those that do not lead to the goal (*distractor items*). Miller and Remington's (2004) study focused on the interaction between the structure of a web site (breadth vs. depth) and the discriminability of target items from distractor items on performance time. Miller and Remington found that when target items were clearly discriminable from the surrounding distractor items, at each page on the path to the goal, then a deeper and less broad structure lead to faster search times. When target items were not clearly discriminable from distractor items, at each page on the path to the goal, then a broader and less deep site structure lead to faster search times.

The sensitivity of web search behaviour to label relevance has been further investigated with analytic techniques that predict search behaviour on the basis of statistics derived from text corpora (Pirolli & Card, 1999; Turney, 2001; Farahat, Pirolli, & Markova, 2004). For example, latent semantic analysis (LSA; Landauer & Dumais, 1997) and point-mutual information (PMI; Turney, 2001; Farahat, Pirolli, & Markova, 2004) have been applied to derive automatic assessments of the relevance of an items label in the context of a given goal statement. These statistical methods have been applied to predict which of the items on a web page people are likely to select for their information goal (Chi et al., 2003). Indeed, Blackmon and colleagues (Blackmon, Polson, Kitajima, & Lewis, 2002; Blackmon, Kitajima, & Polson, 2003) have proposed techniques by which the usability of a web site may be improved by considering the relevance of the labelled links on a page. An implicit assumption made in these tools is that people consider all of the items on a page and select the most relevant one to their goal.

2.2. Strategies for Controlling Interactive Search

While theories, such as information foraging theory (Chi et al., 2003, 2001, 2000; Pirolli & Card, 1999; Pirolli & Fu, 2003), and tools, such as those reported by Blackmon and colleagues (Blackmon et al., 2002, 2003), have assumed that the relevance of all of the links on a page are considered prior to selection, there is no evidence reported to support this assumption. In fact very early work on how people search database menu pages, provided evidence that, while people sometimes assess every possible option prior to selection, they often do not (MacGregor, Lee, & Lam, 1986).

A study by MacGregor, Lee, and Lam (1986), which predated the invention of the web, observed a range of interactive search behaviours, while people search database menu pages. In the experiment, participants searched single-page menus, which differed in terms of the number of items, and whether the participant could see all the menu items at the same time (simultaneous search) or only a single menu item at a time (sequential search). This latter sequential search condition allowed for participants search behaviour to be inferred based on the number of items the participant chose to uncover prior to the selection of an item. MacGregor, Lee, and Lam observed three behaviours, which the author's labelled: *self-terminating*, *exhaustive*, and *redundant*. The self-terminating behaviour consisted of a participant looking at and evaluating each item in turn until one was examined that was considered sufficiently relevant that it was selected immediately. The exhaustive behaviour was evident when people first looked at and evaluated all of the menu items and then returned to and selected the one with the best evaluation. The redundant behaviour consisted of repeatedly looking at and evaluating some subset of the items before making a selection. None of the participants in the study consistently adopted only one of the search behaviours, and two-thirds of the participants showed all three. Furthermore, MacGregor et al. observed that search strategy was contingent on the size of the choice set: The frequency with which each search behaviour was evident was dependent upon the number of items in the menu choice set; in particular, as the number of items increased participants were more likely to self-terminate.

In a similar study, Pierce, Parkinson, and Sisson (1992) considered the implications of label relevance for strategy selection. The experiment used a similar methodology as MacGregor, Lee, and Lam (1986), where participants searched single-page menus in which the semantic relevance of a target item was varied. Pierce, Parkinson, and Sisson found that when the target item was highly relevant to the participants' search goal, participants were more likely to self-terminate by selecting it without assessing any further items. When the target item was less relevant to the goal participants were more likely to exhibit exhaustive or redundant search behaviour.

Studies of people learning computer application menus, rather than searching database menu pages, are also relevant to scoping the range of possible interactive search strategies. Computer application menus are used to access the functional features of a computer application. For example, Microsoft Word has a toolbar menu that contains features for, amongst others, under the "file" option for opening files, saving files etc. Franzke (1994, 1995) and Rieman (1994) observed people as they learnt how to use a novel graphing package (Cricketgraph). They found that participants demonstrated a label following strategy, in which they tended to select items from the application menu with labels that had a high semantic overlap with the current goal. Rieman (1994) gained further understanding of participants' exploratory behaviour by focusing on the search behaviour leading up to the selection of an item. Analyses of verbal protocols and mouse movements suggested that, prior to the selection of an item participants, would often not assess all of the items in the available choice set and would repeatedly reassess an increasingly small subset of those items that were initially assessed. Participants also invested more time on each successive assessment of an item. Rieman, Young, and Howes (1996) later characterized this search strategy as an *iterative deepening of attention*, involving the progressive focusing on a set of potential items with greater effort placed in to thinking about the meaning of an items label on subsequent passes. This search behaviour does not fit into the exhaustive, redundant, or self-terminating taxonomy proposed by MacGregor, Lee, and Lam (1986) for command-menu search. Moreover, the search behaviour observed in the studies by Franzke and Rieman suggest that multiple assessment methods are deployed during interactive search.

The behaviours observed when people search command-menus (MacGregor, Lee, & Lam, 1986; Pierce, Parkinson, & Sisson, 1992) or learn a computer application menu (Franzke, 1994, 1995; Rieman, 1994) suggest different ways in which people may control interactive search. It is an open question whether people adopt similar behaviours during web-based interactive search. First, the content of command-menus and computer application menus is usually substantially different to the content of web pages. Second, studies of command-menu search (MacGregor, Lee, & Lam, 1986; Pierce, Parkinson, & Sisson, 1992) used a sequential presentation methodology to infer participants search behaviour, which can substantially affect strategy selection (Lohse & Johnson, 1996). Studies of application menus (Franzke, 1994, 1995; Rieman, 1994) used a potentially invasive concurrent protocol strategy. These methodological issues are described in more detail in Chapter 3. Moreover, the interest here was to use eye movement data to expose interactive search strategies. It is necessary to first review the substantial theoretical literature because various predictions, which go beyond the available data, concerning these strategies can be derived from existing models.

2.3. Computational Cognitive Models of Interactive Search

There have been a number of models of the cognitive processes that might be involved in controlling interactive search (Cox & Young, 2004; Howes, 1994; Howes, Payne, & Richardson, 2002; Lee & MacGregor, 1985; MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Rieman, Young, & Howes, 1996; Young, 1998). In general, people are assumed to be sensitive to some form of estimate of the likelihood that a labelled item will lead to the goal, however the models differ in, for example, which items are considered, and in the selection strategy that makes use of these estimates. One dimension on which these models can be compared concerns the assumptions that are made about how people choose whether to select an item or continue assessing items (i.e., what people are sensitive to when searching for information).

To account for their empirical findings, MacGregor, Lee, and Lam (1986) described a model of single-page search in which the decision as to whether to select an item, assess a new item, or reassess an existing item was sensitive to the value of the most recently assessed item relative to a threshold. A number of behaviours were emergent from this *threshold-based* strategy. First, if the likelihood that an item would lead to the goal clearly exceeded a selection threshold then there was a chance that the item would be selected immediately, without further evaluation. Second, if an item was only just above the threshold then it would be considered as a possible choice and that evaluation of other items would continue. Finally, if after having examined all of the items, more than one just exceeded the threshold then the model re-examined this subset. The three empirically observed search behaviours, described earlier—self-terminating, exhaustive, and redundant—were emergent from the assumption that people used a threshold-based strategy.

Like MacGregor, Lee, and Lam (1986), Miller and Remington (2004) assumed that people were sensitive to a selection threshold; however they extended their analysis to the search of a multi-page web site. In Miller and Remington's model, items on the current page were assessed, and if an item exceeded a threshold it was selected. If when all the items on the current page were assessed, none exceeded the threshold, then the threshold was lowered and the items on that page re-evaluated relative to the new, lowered threshold. The model backed up to the previous page in the site when none of the items on the current page exceed the reduced threshold.

Howes, Richardson, and Payne (2002) reported a model in which the decision as to whether to leave a page (i.e., to select a backup button) was moderated not only by the relevance of the items on the current page, but also by the relevance of items on the previously visited pages of the current site. Howes et al.'s model was also a threshold model, but importantly, one in which the threshold was dynamically determined by the distal search context (i.e., memory for items that were not necessarily available on the current menu page). Consistent with the empirical data (Payne, Richardson, & Howes, 2000), Howes et al.'s model used an episodic memory of previous assessments to determine whether the utility of backing up was greater than the utility of an available forward move.

Rieman, Young, and Howes (1996) proposed a model, which was called Iteratively Deepening Exploratory Search (IDXL). The model captured Rieman's earlier observation that multiple assessment methods are deployed during interactive search. The model searched both an external menu and the internal space of possible evaluations and was sensitive to the costs and benefits of different methods of assessing items. The model evaluated menu items in turn, starting with a relatively low cost evaluation of the menu items and moving to a more sophisticated, but higher cost assessment procedure. A low cost assessment might be characterized by "*Does the currently attended item contain a word that is also in the explicitly articulated goal description?*" Appling this assessment procedure sometimes identified items that provided exact label matches with the goal description, which resulted in the selection of an item. If none of the items provided an exact label match with the goal description, then the model would reassess a subset of the menu items with a higher cost, but more sophisticated assessment procedure, such as "*Is there a semantic link between an items label and the goal?*" The model exhibited behaviours consistent with observations of participant learning computer applications menus (Franzke, 1994, 1995; Rieman, 1994): exact label matches were selected sooner than labels that were synonyms of the goal description, and the model repeated scanning of a subset of the available menu items, with increasing attention to items on each successive pass. Moreover, IDXL went beyond MacGregor, Lee, and Lam's (1986) model by embedding hypotheses about the details of the cognitive processing that is conducted during interactive search.

It worth noting that a common feature of the accounts of interactive search described above (e.g., MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Reiman, Young, & Howes, 1996) is that they assume that the decision to select is sensitive to the value of the most recently assessed item relative to a threshold. Such threshold-based accounts bear similarity to Simon's (1969) notion of satisficing: "Aspiration levels provide a computational mechanism for satisficing. An alternative satisfices if it meets aspirations along all dimensions. If no such alternative is found, search is undertaken for new alternatives. Meanwhile, aspirations along one or more dimensions drift down gradually until a satisfactory new alternative is found or some existing alternative satisfices. A theory of choice employing these mechanisms acknowledges the limits on human computation and fits our empirical observations of human decision making far better than the utility maximisation theory" (Simon, 1969, pp. 30).

In contrast, Young (1998), and more recently Cox and Young (2004), have proposed a rational account (Anderson, 1990) of interactive search where the assessment of items continues until the cost, in terms of the time required to perform the assessment, outweighs the estimate of the information to be gained from further assessment. In this account, the value of assessment is dependent on the context provided by all previous assessments of items in the current choice set. Following Rieman, Young, and Howes (1996), Cox and Young assumed that multiple assessment methods of varying quality and cost could be applied to a given menu item to provide an independent assessment of an items' relevance, reflecting a subjective judgment of the likelihood that the selection of the item will lead to the goal. Importantly, the model also assumes that only a single item in the menu choice set would lead to the achievement of the goal (i.e., that there is only a single target item and that the remaining items in the menu are all distractors). Taken in this context, the utility of assessing an item can be evaluated by the expected information gain in reducing the degree of uncertainty as to which of the items in the menu choice set actually leads to the goal. In other words, the model utilizes the landscape of the relevance values *across all the items in the menu choice set*. Assessments of items are favoured that are expected to further reduce this uncertainty. The model opts to select the item from the menu with the greatest relevance estimate when the expected reduction in uncertainty (i.e., the information gained) from further assessment is no longer worth the cost incurred. Moreover, the model assumes that people will exhibit a broad range of search behaviours that are emergent from a single decision strategy that takes into account the context of the set of assessments made so far.

A key assumption in Cox and Young's work (Cox & Young, 2004; Young & Cox, 2000; Young, 1998) is that a subjective assessment of the likelihood that a given menu option will lead to the goal is dependent on other assessments made. More specifically, Cox and Young assume that the information gained (i.e., reduction in uncertainty as to which option is correct) by assessing the relevance of a menu option depends on the relevance of the other options in the choice set. This *normalization assumption* "reflects real cross-relationships between the judgments about choices made by a person, and cannot be avoided … the reality is that people are often forced to make rapid and radical revisions of their estimates of the correctness of particular options as they work their way through [the options available]" (Young, 1998, pp. 474).

Cox and Young's (2004) model is particularly interesting because it makes a novel and empirically untested prediction. Hitherto a general assumption that has been made in the literature (e.g., Blackmon et al., 2002, 2003; Chi et al., 2003; Miller & Remington, 2004; Pirolli & Fu, 2003) is that estimates of an item's relevance, which determine the subjective value of selecting the item, are independent of context. Cox and Young's framework, however, suggests that not only the relevance of an item affects the decision of whether to select it, but that the remaining distractor items in the choice set also influence this decision. In other words, that the subjective value of selecting an item is sensitive to the context provided by the previously visited item in the choice set.

Information foraging theory (Pirolli & Card, 1999; Pirolli & Fu, 2003; Pirolli, 2005; Pirolli, in press) assumes that during web-based information gathering activities, people are sensitive to the rate of gain of information per unit cost. A key contribution of this work is the hypothesis that people will leave a site/page when the rate of gaining information falls below the average rate of gain. One of the assumptions of Pirolli and Fu's (2003) model, SNIF-ACT, is that people consider the likelihood of every item on a page and therefore, while site-leaving decisions are sensitive to the rate of information gain, the choice of which items to assess is not. Moreover, an *assess-all* decision strategy is used in SNIF-ACT to control search of each page. This simplification is potentially non-trivial, if as is suggested by MacGregor, Lee, and Lam's (1986) command-menu data, people sometimes choose to select an item without assessing any further items in the choice set (i.e., self-terminating search).

2.4. Summary

Previous empirical studies of interactive search have focused on regularities in how people search web pages (e.g., Card et al., 2001) and also database menu pages (e.g., MacGregor, Lee, & Lam, 1986; Pierce, Parkinson, & Sisson, 1992). Unsurprisingly, a number of studies have found that people tend to select items that are more relevant to their goal (Blackmon, Polson, Kitajima, & Lewis, 2002; Blackmon, Kitajima, & Polson, 2003; Card et al., 2001; Chi et al., 2003, 2001, 2000; Church & Keane, 2004; Katz & Byrne, 2003; Kaur & Hornof, 2005; Miller & Remington, 2004; Pierce, Parkinson, & Sisson, 1992; Pirolli & Fu, 2003). The ease by which items that actually lead to the goal (*target items*) can be readably discriminated from those that do not lead to the goal (*distractor items*) has been found to influence participants search behaviour (Miller & Remington, 2004; Pierce, Parkinson, & Sisson, 1992).

While there is somewhat limited empirical evidence concerning how search behaviour might be controlled, there is a substantial theoretical literature in which a number of cognitive models of interactive search have been proposed (Cox & Young, 2004; Howes, Payne, & Richardson, 2002; MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Rieman, Young, & Howes, 1996; Young, 1998). Importantly these models differ in terms of the assumed strategy for determining when the selection of an item should occur during interactive search. First, *assess-all* accounts assume that people consider all of the items on a page prior to making a selection (i.e., take the best, without a selection threshold), as is exemplified in Pirolli and Card (1999). Second, *simple threshold* accounts assume that people make a selection immediately following an assessment of a highly relevant item, as is exemplified in Miller and Remington (2004). Finally, *context-sensitive* accounts assume that the decision as to whether to select an item might be sensitive to the entire set of assessments made so far (Cox & Young, 2004; Young, 1998), and not just to the most recent item (i.e., a context determined dynamic selection threshold).

CHAPTER 3: THE EFFECT OF RELEVANCE OF SELECTION

The aim of Chapter 3 is to discriminate between the different hypotheses concerning strategies for determining when selection of an item should occur during interactive search. As defined in the previous chapter, models of interactive search have assumed an assess-all, threshold or context-sensitive selection strategy. A series of experiments are reported which consider how systematically manipulating the relevance of the items in the local choice set influenced the participant's decision to select an item.

3.1. Eye-tracking Methodology

The theoretical motivation for the current study was to distinguish between different accounts of interactive search. It was, therefore, imperative to identify which of the set of items participants chose to assess in order to discriminate between different accounts of interactive search. Previous studies of the strategies that people use to search database menu pages (MacGregor, Lee, & Lam, 1986; Pierce, Parkinson, & Sisson, 1992) used a process tracing methodology in which menu alternatives were at first hidden and then exposed one-by-one whenever a down-arrow key was pressed. Information acquisition behaviour, however, is influenced by the cost of accessing information from the environment (Lohse & Johnson, 1996). It could be the case that search behaviours observed by MacGregor, Lee and Lam (1986)–self-terminating, exhaustive, and redundant–were a reflection of the cost structure imposed by the process-tracing methodology.

A solution to this problem is to infer participants search strategy from analysis of eye movement protocols. Eye movement protocols provide moment-to-moment behavioural index of users' human-computer interactions. Eye movements provide an on-line indication of how people acquire and process information, and have provided significant benefits in the analysis of cognitive processes in a variety of task domains, such as reading (Just & Carpenter, 1980, 1984; Schilling, Rayner, & Chumbley, 1998), equation solving (Salvucci & Anderson, 2001), menu selection (Aaltonen, Hyrskykari, & Räihä, 1998; Byrne, Anderson, Douglass, & Matessa, 1999; Hornof, 2004), and web search (Card et al., 2001).

In general, during normal view conditions, the eye will alternate between periods of rapid movement of the eye (saccade), and stationary periods where constant gaze is maintained (fixations). It is generally accepted that visual information is available only during fixations, and not during saccades (Findlay & Gilchrist, 2003). The active vision approach (Findlay & Gilchrist, 2003; Just & Carpenter, 1984; Liversedge & Findlay, 2000) assumes that eye movement fixations and gaze shifts are tightly coupled with the allocation of visual perception and cognition. Nonetheless, caution is required when interpreting the assumed relationship between eye-movements and cognitive processes (see Anderson, Bothell, & Douglass, 2004). For instance, eye movement fixations can be used to infer information acquisition from the environment, but cannot be used to infer, for instance, memory retrieval processes.

In order to gain detailed and accurate insights into participants' interactive search behaviour, eye movement protocols were analysed as a primary dependent variable in all of the empirical studies presented in this thesis. Throughout this thesis, I adopt the convention found in the reading literature (e.g., Rayner & Pollatsek, 1989) where multiple, immediately successive fixations to an item, are aggregated to an *item-gaze*. I refer to an item-gaze as a *visit*. I assume that an eye movement gaze directed towards an item in the menu can be broadly mapped to the cognitive process of making an assessment of the probability that an item will lead to the goal. It is important to note at the outset that the main conclusions of this thesis are not dependent on this assumption (see Chapter 6 for details).

3.2. Experiment 1

The first empirical study was designed to provide evidence to distinguish between different accounts of interactive search by systematically manipulating the relevance of the distractor labels and measuring the consequences for which of the set of items are assessed. As the mean relevance of the set increases, assess-all accounts (e.g., Pirolli & Card, 1999; Pirolli & Fu 2003) predict no change because all items are assessed regardless. Threshold accounts (e.g., Miller & Remington, 2004) predict that fewer items will be assessed on the first pass as the average relevance of the distractor items increases, because on average an item will be more likely to exceed the threshold earlier. Context-sensitive accounts (e.g., Cox & Young, 2004; Young, 1998) predict that people will assess more items as the average relevance of the distractor items increases, because the relative value of further assessment will be greater when there are many distractor items that are relevant to the goal.

Cox and Young’s (2004) model also predicts that revisits to items will be common, reflecting the application of assessment methods, of varying cost and quality, to items prior to selection. In this respect, it is consistent with the related empirical (Franzke, 1994, 1995; Rieman, 1994) and theoretical (Rieman, Young, & Howes, 1996) work concerning how people learn to use a novel computer application interface through exploratory learning. In particular, participants should be more likely to revisit items that are more relevant to the goal description, and to place greater effort (i.e., more time) in thinking about the meaning of an items label on subsequent revisits. The models proposed by Pirolli and Card (1999), Pirolli and Fu (2003), and Miller and Remington (2004) do not make these predictions. Assess-all accounts assume that participants visit all of the items in a choice set prior to selection, and make very few revisits to items (revisits would only occur for relocating a previously visited item for selection). In the basic threshold model very few revisits are expected. In Miller and Remington’s threshold lowering model, revisits are made but they are not guided by semantic relevance.

Moreover, it is important to note that these predictions do not concern the number of fixations that will be made but the number of items that will be fixated and the duration of each block of consecutive fixations to a particular item.

In Experiment 1, the relevance of the labelled items was manipulated. In order to put together menu choice sets for the experiment, the relevance of the sampled web-page labels for each search goal from a particular category was determined. A number of automated tools are available to compute the semantic similarity between a label and a goal description, most notably Latent Semantic Analysis (LSA; Landauer & Dumais, 1997, available at http://lsa.colorado.edu/).

Blackmon and colleagues (Blackmon, Kitajima, & Polson, 2005, 2003) have explicitly outlined how LSA might be applied to estimating the degree of semantic similarity between a user's goal and a link label on a web page. LSA is a machine learning technique that builds a semantic space reflecting the statistical properties of the linguistic environment. This semantic space is learnt by applying singular value decomposition, a mathematical procedure similar to factor analysis, to a training corpus. A given entry is represented as a vector in a high dimensional semantic space. The similarity between any pair of texts can be defined by the cosine value between the corresponding two vectors. Each cosine value is between +1 and -1, where values closure to +1 indicates that two snippets of text are more semantically similar, and values closure to -1 indicates that two snippets of text are more semantically dissimilar.

3.2.1. Method

3.2.1.1. Participants

Sixteen Cardiff University undergraduate psychology students participated in return for course-related credit. All participants were native English speakers and had normal uncorrected vision. All participants were experienced in using a World Wide Web browser, and all had been required to use various computer software packages to produce coursework.

3.2.1.2. Materials

Participants completed ecologically valid interactive search tasks which required them to search a simplified web page (or menu) for information relevant to a given goal statement. In order to derive ecologically determined goal statements, a web usage survey was posted to under-graduate students in the School of Psychology at Cardiff University. The web usage survey aimed to identify search queries, which participants in the experiment would typical have used the web for. From the 25 responses to the survey (approximately 5% response rate) it was possible to determine 45 unique search goals. Search goals were then split into 12 broad category types, (e.g., news, retail and shopping, banking, university-related etc). The web usage survey also aimed to find out which web sites the respondents had visited whilst searching the web for each goal. The labels from these suggested web sites were then sampled and collected together under one of the 12 category types outlined above. For example, for the search goal "*check your bank balance*" labels were sampled from various online banking web sites (e.g., http://www.hsbc.co.uk & http://www.natwest.co.uk). In total, some 2,000 individual labels were sample from various web sites for all category types.

Table 3.1. Mapping LSA cosine scores to discrete values of label relevance

Transformed value	LSA Cosine Value
1	> .418 and =< .490
2	> .346 and =< .418
3	> .274 and =< .346
4	> .202 and =< .274
5	> .144 and =< .202

Experiment 1 used LSA as an automated tool to derive independent estimates of the relevance of each of the sampled web-page labels in relation to a particular goal statement. The *general reading up to 1st year of collage (300 factors)* database was used (available at: http://lsa.colorado.edu/). In order to determine labels of differing relevance, the distribution of returned LSA cosine values were portioned into discrete boundaries. More specifically, LSA cosine values were transformed to a 5-point scale, defined as equally spaced regions between the range (*M* = .286, *SD* = .144, *min* = .13 and *max* = .49), where 1 represented a label that was very relevant to the goal description, and 5 represented a label that was not at all relevant to the goal description (see Table 3.1).

For each goal statement there was a collection of labels, each of which had a relevance rating derived using LSA. In order to devise 16-item menus for each of the goal statements, a single independent judge was recruited. The judge was instructed to randomly select labels for each search goal to generate menus for each of the experimental conditions. For each menu, regardless of experimental condition, a single target item was identified (defined by a transformed score = 1). For the distractor items, 10 labels were identified which were moderately relevant to the goal statement (defined by a transformed score = 3) and 15 labels were identified which were not relevant to the goal statement (defined by a transformed score = 5, but could also = 4). This arrangement allowed each goal statement to be placed into any of the experimental conditions

3.2.1.3. Design

In the current study, the number of moderately relevant distractor item's that were present in the menu was manipulated as a within-subjects factor. The relevance of an items label for a given goal was determined using LSA (described in more detail in the next section). Each menu contained 16-items and contained only a single target item that was scored as highly relevant to the search goal. Menus were devised such that for each condition participants searched menus that differed in terms of the semantic relevance of the remaining distractor (or non-target) items. By default, the remaining distractors in the menu were scored as being not relevant to the search goal, and the number of distractors scored as moderately relevant to the goal being varied between conditions. The number of moderately relevant distractors in each menu was either: none, two, five, and ten. The primary measure examined in the study was eye-tracking data of participants' eye movements up to and including the first selection of an item.

3.2.1.4. Procedure

In the experiment, participants completed 20 interactive search tasks (or trials). For each trial the participant was required to search a simplified web page (or menu) for information relevant to a given goal statement. There were five trials for each of the experimental conditions. The design was counter-balanced, such that across different participants in the study, a given goal statement was placed in different experimental conditions. The experimental materials were controlled by a purpose-built Microsoft Visual Basic program presented on a high contrast 19 inch FC Trinitron CRT monitor. All menus contained 16 labelled links, of which only one led to the completion of the goal (i.e., one target item and 15 distractor items). The items in each menu were presented in a standardized format: characters were font 15 Comic Sans MS and labels were presented in a single vertical list with an approximate distance between each label of three degrees of visual angle. The target item was always located towards the top of the list of menu (between menu positions 3 through 8). Each trial commenced by the participant first reading the goal statement. When the participant was ready, they selected a search button with the mouse, which then removed the goal statement and displayed the menu on the screen. Participants were instructed to scan through the menu, commencing their search at the top of the menu, and to select the item that they believed to be most relevant to the goal statement as quickly and accurately as possible. In order to impose a meaningful cost structure to the task, participants did not progress to the next trial until they correctly selected the single target item from each menu. Eye tracking was performed using an ASL Pan/Tilt optics eye tracking system. Eye movement data was sampled at a rate of 50 times per second (once every 20 ms). Eye movement fixations were determined using the Applied Science Laboratories *Eyenal* software package.

3.2.2. Results

For all experimental analyses reported throughout this thesis, I was only interested in participants' interactive search behaviour up to the initial selection of an item from the menu for a given trial. A particular focus was analysis of eye movement protocols. Figure 3.1 presents a typical eye movement trace from Experiment 1. Fixations were mapped to an item in the menu, if they landed within the item's respective area of interest. Areas of interest were defined as a standardized rectangular area around each menu item (occurring at the mid-point between vertically contiguous items). Fixations that did not land over a menu item were ignored (accounting for less than 5% of all fixations).

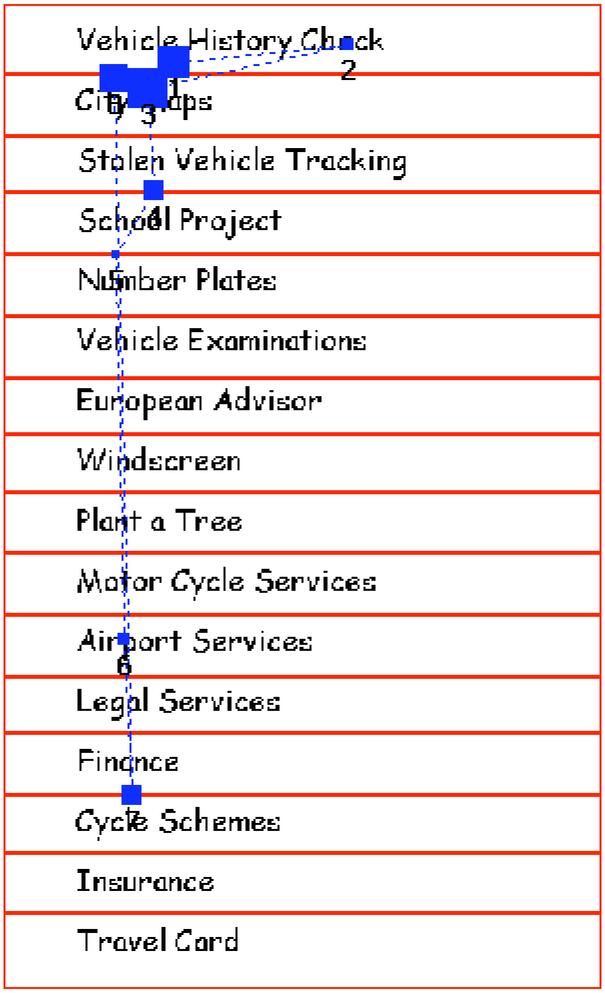


Figure 3.1. **A typical eye movement trace, where the goal statement for this menu was "Find a road map of Cardiff", and the second item in the menu "City Maps" was the target. Rectangular boxes around menu items define areas of interest.**

Table 3.2. *Results for Experiment 1 for the main dependent variables*

	Number of moderate distractor items										
	None-moderate		Two-moderate		Five-moderate		Ten-Moderate				
Dependent Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*	*MSE*
Accuracy	78.75%	17.08%	83.75%	16.68%	80.00%	16.12%	82.50%	16.12%	.449	.719	.019
Time required to selection	7.49s	2.25s	7.51s	1.91s	8.67s	3.62s	7.59s	2.83s	1.068	.372	4.233
Number of items visited at least once	11.44 items	1.50 items	10.69 items	1.34 items	11.25 items	2.53 items	10.37 items	2.00 items	2.358	.145	2.379

Table 3.2 presents the main dependent variables from Experiment 1. For all statistical analysis, a one-way repeated measures analysis of variance (ANOVA) was employed, adopting an alpha-level of $p < .05$ for statistical significance. All trials were analysed regardless of correctness of participants' initial selection. The main dependent variables were: 1) Accuracy, which was the proportion of trials in which the target item correctly selected on the first selection; 2) Time required to selection, which was the time from when the participant first started searching the menu to the first selection of an item; 3) Number of items visited at least once, which was the number of items in the menu that were fixated upon at least once prior to the selection of an item.

It was found that the manipulation of distractor relevance did not have a significant effect on any of the main dependent variables (see Table 3.2). In particular, participants' visited approximately the same number of items in the menu prior to selection, regardless of the number of distractor items in the menu that were moderately relevant to the goal statement.

3.2.3. Discussion

Experiment 1 was designed to provide evidence to distinguish between different accounts of interactive search by systematically manipulating the relevance of the items presented in a menu choice set. Ecologically-valid goal statements were determined and for each a collection of web-page labels sampled. Latent Semantic Analysis (LSA) was used as a tool to estimate the semantic relevance of a label for a given goal statement. For the experimental conditions, an independent judge devised menus that contained 16 labelled items and varied in terms of semantic relevance to the goal statement. Analysis of performance and eye-movement protocols revealed no significant affect of label relevance on participants search behaviour.

Experiment 1 used LSA as an automated tool to derive independent estimates of the relevance of each of the sampled web-page labels in relation to a particular goal statement. LSA is a statistical machine learning technique which has shown to accurately model human semantics judgments over large data sets (e.g., Landauer & Dumais, 1997; Landauer, 1998) and clearly its use in this context could provide a degree of independent objectivity in defining label relevance. It was clear, however, that some erroneous judgments of label relevance had been made by LSA. Some of these were attributable to the briefness of the description of the goal statement and/or label that did not provide sufficient content for LSA to compute an accurate similarity score (see Blackmon et al., 2003).

Furthermore, scale was also an important concern in the current context. Machine learning techniques, such as LSA, have been shown to be reliable over large data sets (e.g., Landauer & Dumais, 1997; Landauer, 1998). For most of the experiments reported throughout this thesis, however, the number of trials per condition was relatively small (approx. 4 – 6 trials). This was mainly due to time constraints imposed by the difficulties inherent in using eye-tracking methodology. In retrospect, given that the sample size for the set of labels actually used was relatively small (totalling some 560 labels) it would be surprising if LSA could provide a highly accurate method for discriminating label relevance. Consequently, for the remaining experiments presented throughout this thesis label relevance was defined by human relevance judgments. Human relevance judgements provide a more conservative method to estimating label relevance (e.g., Lesk & Salton, 1968). It is of course further debatable whether these should be considered as a gold standard for determining semantic relevance.

3.3. Experiment 2

Experiment 2 replicated the initial experiment, with the exception that human judges rated labels, and that both the relevance of the target and the distractor items in the menu were systematically manipulated. The aim of the study was to distinguish between different accounts of interactive search by manipulating the relevance of the items in the local choice set and observing the consequences for which of the set of items were assessed. The predictions for Experiment 2 are, therefore, the same as those described in the introduction to Experiment1.

3.3.1. Method

3.3.1.1. Participants

Twenty Cardiff University undergraduate psychology students participated in return for course-related credit. None of the participants had taken part in any of the other experiments reported in this thesis. All participants were native English speakers and had normal uncorrected vision. All participants were experienced in using a World Wide Web browser, and all had been required to use various computer software packages to produce coursework.

3.3.1.2. Materials

In order to put together menu choice sets for the experiment, the relevance of each of the sampled web-page labels, in relation to a particular goal statement, had to be first determined. Thirteen Cardiff University undergraduate psychology students completed a ratings questionnaire in return for course-related credit. (None of the participants that took part in the rating study took part in any of the subsequent experiments.) Participants were instructed to estimate the likelihood that selecting a label would lead to the achievement of the goal. Participants made relevance estimates on a 5-point scale, where 1 represented a label that was very relevant to the goal description and 5 represented a label that was not at all relevant to the goal description. To gather the ratings all of the sampled web-labels for a particular goal description were made available at once and participants asked to rate them one-by-one. Participants were instructed to make each relevance estimate *independently* of the estimate of the relevance of the other labels presented on the page. Based on the ratings of the sampled web-page labels, it was possible to construct 16-item menus with labels of varying semantic relevance to a particular goal statement.

3.3.1.3. Design

The experiment was a within-subjects design with two independent variables. Participants searched menus that differed in terms of the semantic relevance of the target item (highly relevant or moderately relevant) and the semantic relevance of the distractor items (moderately relevant, not relevant, or not at all relevant). The relevance of an items label for a given goal was determined from ratings provided by a separate group of participants that did not take part in the menu search experiment (see Materials section below for more details). Importantly, ratings of a labels' relevance were made on a 5-point scale, where 1 represented a label that was very relevant to the goal description, and 5 represented a label that was not at all relevant to the goal description. Consequently, a highly relevant target item was defined as a label that received a median rating of 1 and a moderately relevant target item received a median rating of 2. As for the distractors, each condition differed in the average relevance of all of the remaining items in the choice set: In the moderate distractor condition, the mean rating of the items was 3; for the poor distractor condition the mean rating of the items was 4; for the very poor distractor condition the mean rating of the items was 5. The primary measure examined in the study was eye-tracking data of participants' eye movements up to and including the first selection of an item.

3.3.1.4. Procedure

In the current study, participants completed 40 interactive search tasks that required them to search a simplified web page (or menu) for information relevant to a given goal statement. The goal statements were the same as those generated from the web-usage survey posted to under-graduates for Experiment 1. The semantic quality of each of the sampled web-page labels in relation to a particular goal statement was determined from human relevance judgments. Based on these ratings of the sampled web-page labels, it was possible to construct 16-item menus with labels of varying semantic relevance to a particular goal statement. Participants searched menus that differed in terms of the semantic relevance of the target item (highly relevant or moderately relevant) and the semantic relevance of the distractor items (moderately relevant, not relevant, or not at all relevant).

The experimental materials were controlled by a purpose-built Microsoft Visual Basic program presented on a high contrast 19 inch FC Trinitron CRT monitor. The menu items were presented in a vertical list in font 15 Comic Sans MS and the approximate distance between each label was three degrees of visual angle. Each trial commenced by the participant first reading the goal statement. When the participant was ready, they selected a search button with the mouse, which then removed the goal statement and displayed the menu on the screen. Participants were instructed to scan through the menu items, commencing their search at the top of the menu, and to select the item that they believed to be most relevant to the goal statement as quickly and accurately as possible. As before, in order to impose a meaningful cost structure to the task, participants did not progress to the next trial until they selected the target item. The target item was always located towards the top of the list of menu (between menu positions 3 through 8). Eye movement data was recording using an ASL Pan/Tilt optics eye tracking system, which was sampled at a rate of 50 times per second. Eye movement fixations were determined using the same procedure outlined in Experiment 1.

3.3.2. Results

As in Experiment 1, for each trial I was only interested in participants' search behaviour from the beginning of a trial up to the initial selection of an item. This included analysis of eye movement protocols. All trials were analyzed regardless of correctness of participants' initial selection.

3.3.2.1. Time required to and accuracy of initial selection

A set of analyses of the effects of semantic relevance on selection was conducted. These tests were not the primary aim of the study but provide a background picture of performance. Figure 3.2 presents the average time spent from the start of a trial up to when the participant choose to first select an item. A 2 x 3 (target-relevance x distractor relevance) repeated-measures ANOVA found that participants took less time to select an item when it was highly relevant to the goal compared to when it was moderately relevant to the goal, *F* (1, 19) = 214.01, *p* = .001, *MSE* = 2.92. The relevance of the distractor items also affected the time to select an item, *F* (2, 38) = 8.63, *p* = .001, *MSE* = 4.40. There was a significant linear trend suggesting that participants required significantly more time to select an item when the distractors were moderately relevant, compared to not relevant or not at all relevant, *F* (1, 19) = 17.98, *p* = .001, *MSE* = 4.15. The interaction between target and distractor relevance on time to initial selection was significant, *F* (2, 38) = 16.69, *p* = .001, *MSE* = 2.33. On further analysis of the simple main effects the relevance of the distractors had less of an affect on selection time when the target was moderately relevant compared to when the target was highly relevant to the goal statement, *F* (2, 18) = 3.72, *p* = .044, *F* (2, 18) = 22.69, *p* = .001, respectively.

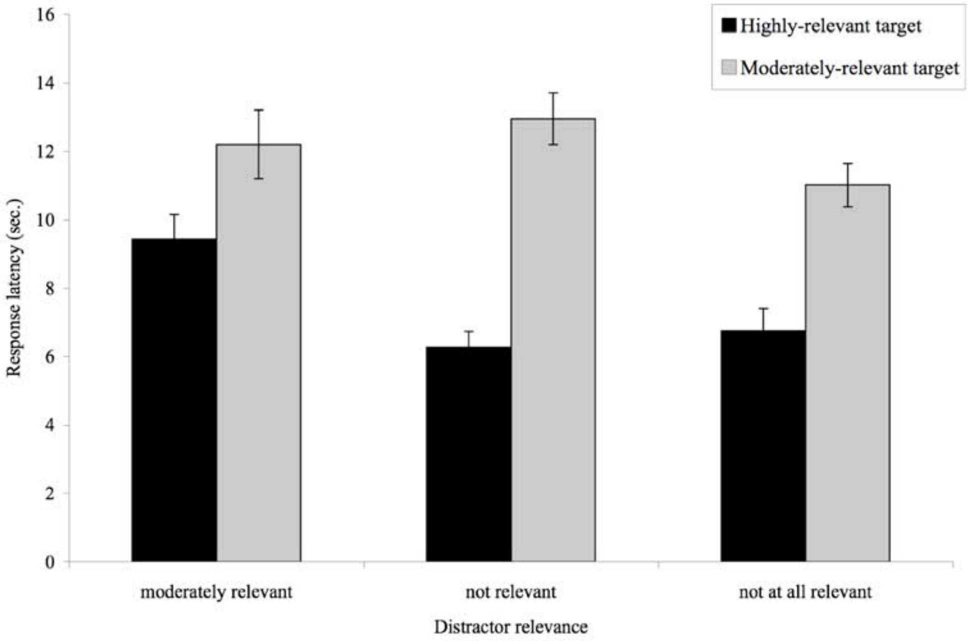


Figure 3.2. Time to the initial selection of an item for Experiment 1.

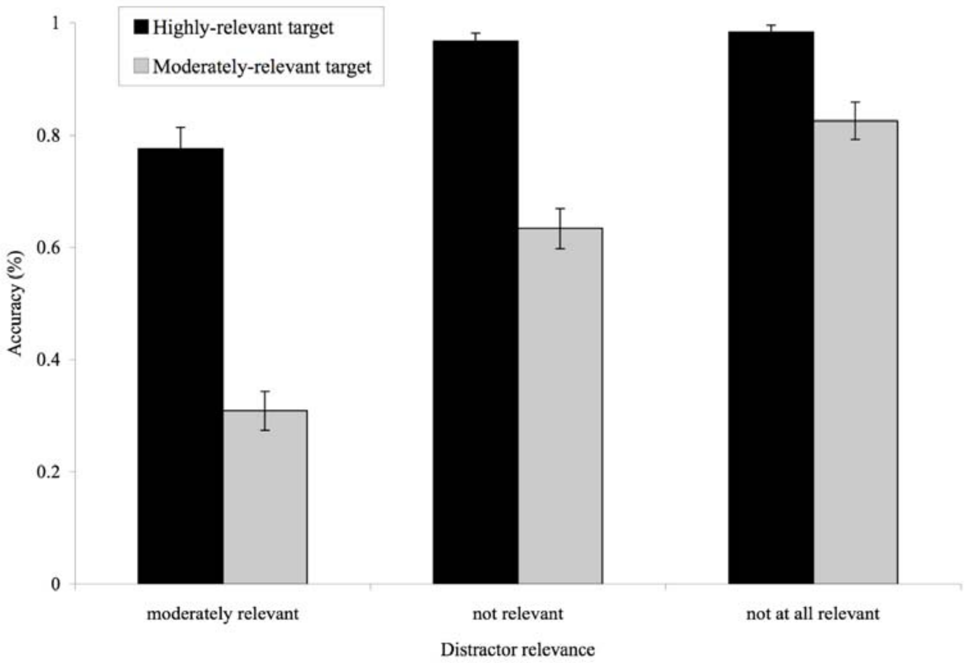


Figure 3.3. Accuracy of initial selection for Experiment 1.

Figure 3.3 presents the average accuracy of participants' initial selection. A 2 x 3 (target-relevance x distractor relevance) repeated-measures ANOVA found that participants were more likely to accurately select the target item when it was highly relevant to the goal compared to when it was moderately relevant to the goal, *F* (1, 19) = 161.59, *p* < .001, *MSE* = .02. The relevance of the distractor items also affected selection accuracy, *F* (2, 38) = 119.00, *p* < .001, *MSE* = .01. There was a significant linear trend suggesting that selection accuracy significantly decreased when the distractors were moderately relevant, compared to not relevant or not at all relevant, *F* (1, 19) = 196.73, *p* < .001, *MSE* = .01. The interaction between target and distractor relevance on selection accuracy was significant, *F* (2, 38) = 9.89, *p* < .001, *MSE* = .02. Further analysis of the simple main effects did not alter the main pattern of results: There was a significant affect of distractor relevance on selection accuracy regardless of whether the target was moderately relevant or highly relevant to the goal statement, *F* (2, 18) = 65.03, *p* < .001, *F* (2, 18) = 14.24, *p* < .001, respectively.

Moreover, analyses of time to selection and selection accuracy indicated that participants were quicker and more likely to select items that were judged highly relevant to the goal statement. It was also found that the relevance of the remaining distractor items in the menu influenced the time taken to the initial selection of an item and the likelihood that this selection was the correct target item. I next provide an analysis of eye movement protocols.

3.3.2.2. Number of items visited/revisited

Figure 3.4 shows the number of items that were visited at least once and the number of items that were revisited (i.e., the number of items that were visited at least twice) for each experimental condition. A 2 x 3 (target relevance x distractor relevance) repeated-measures ANOVA found that the number of items that were visited at least once was significantly affected by the relevance of the target item, *F* (1, 19) = 73.94, *p* < .001, *MSE* = 2.19 and the relevance of the distractor items, *F* (2, 38) = 6.94, *p* < .005, *MSE* = 1.71. The interaction between target and distractor relevance was also significant, *F* (2, 38) = 6.61, *p* = .005, *MSE* = 1.87. On further analysis of the simple main effects found that when the target item was highly relevant then there was a significant affect of distractor relevance on the number of items visited, *F* (2, 18) = 8.78, *p* = .005. But there was no effect of distractor relevance when the target item was moderately relevant to the goal statement, *F* (2, 18) = 2.12, *p* = .149.

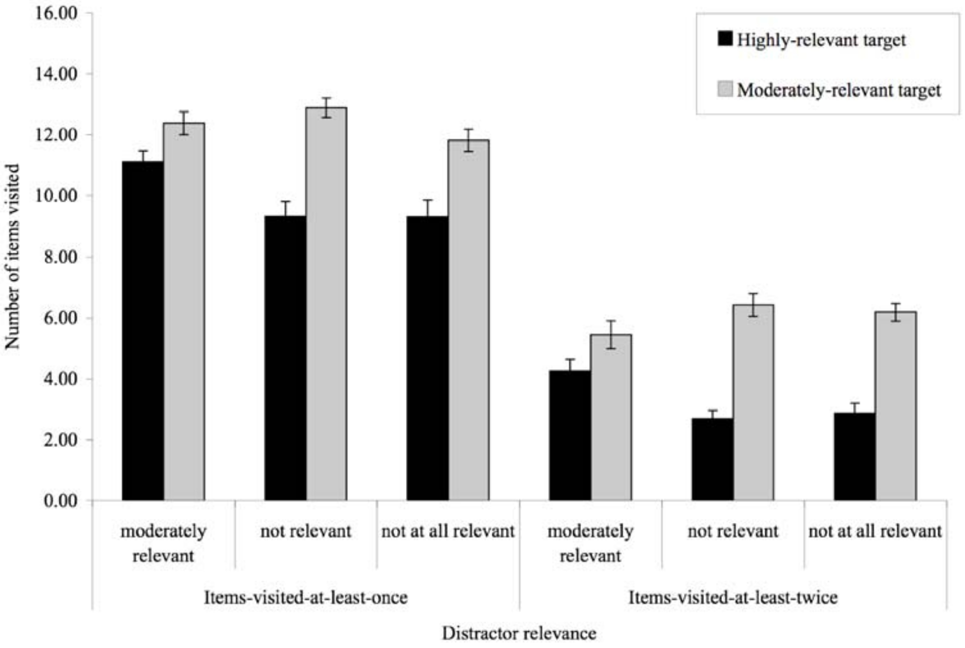


Figure 3.4. The mean number of items visited and revisited up to the initial selection of an item in Experiment 1.

Figure 3.4 shows that revisits to items were common. A 2 x 3 (target relevance x distractor relevance) repeated-measures ANOVA found that the number of items that were revisited was significantly affected by the relevance of the target item, *F* (1, 19) = 226.38, *p* < .001, *MSE* = 1.00. There was no significant main effect of distractor relevance, *F* (2, 38) = .76, *p* = .477, MSE = 1.33. The interaction between target and distractor relevance was significant, *F* (2, 38) = 18.34, *p* < .001, *MSE* = 1.02. On further analysis of the simple main effects found that when the target item was highly relevant then there was a significant affect of distractor relevance on the number of items revisited, *F* (2, 18) = 12.58, *p* = .001. But again, there was no effect of distractor relevance when the target item was moderately relevant to the goal statement, *F* (2, 18) = 3.06, *p* = .072.

Moreover, when the target item was highly relevant then fewer distractor items were visited. In other words, the participants decided to select an item without sampling all the items in the menu. The less relevant the distractors were – the fewer were visited.

3.3.2.3. Duration per visit

A claim to the iterative deepening account is that participants should spend more time on each successive revisit to an item. Consequently, whether the duration of an item gaze increased between an initial visit and subsequent revisits was considered. Participants spent more time on the first visit (*M* = 461 ms, *SD* = 81 ms), compared to the second (*M* = 406 ms, *SD* = 99 ms) or the third (*M* = 405 ms, *SD* = 156 ms) visit to an item. A 3 x 2 x 3 (number of visits x target relevance x distractor relevance) repeated-measures ANOVA found that the duration of an item visit was significantly affected by the number of previous visits to the item, *F* (2, 28) = 5.92, *p* = .007, *MSE* = .018. Tests of within-subjects contrasts found a significant linear trend suggesting that the duration of time spent looking at an item increased over consecutive revisits to the same item, *F* (1, 14) = 7.479, *p* = .05, *MSE* = .022.

The average duration of all item visits was longer when the target item was highly relevant to the goal statement (*M* = 439 ms, *SD* = 102 ms) compared to when the target item was moderately relevant (*M* = 409 ms, *SD* = 135 ms), *F* (1, 14) = 8.19, *p* = .013, *MSE* = .01. (This apparently counterintuitive finding may just reflect that more revisits, which were shorter in duration, were made when the target was moderately relevant.) There was no significant affect of distractor relevance on the duration of an item visit, *F* (2, 28) = 1.262, *p* = .299, *MSE* = .01. All second-order and third-order interactions were also non-significant (for brevity these are not reported).

3.3.2.4. Proportion of first-visit selections

Figure 3.5 shows the distribution of the frequency with which each number of items was visited. The distribution is bimodal. This suggests that participants either chose to select an item after visiting for the first time, a behaviour I refer to as *first-visit-selection* (which is equivalent to MacGregor, Lee and Lam's (1986) description of self-terminating search), or they continued to visit most of the remaining items in the menu.[1]

I next considered, the effect of target and distractor relevance on the proportion of searches where a first-visit-selection was made. Participants were more likely to select an item after visiting for the first time when the target item was highly relevant (*M* = 25.83%, *SD* = 22.43%) compared to moderately relevant (*M* = 8.61%, *SD* = 11.27%). A 2 x 3 (target-relevance x distractor relevance) repeated-measures ANOVA found a significant main effect of target relevance, *F* (1, 19) = 27.71, *p* < 0.01, *MSE* = .03. Participants were also less likely to select an item after visiting it for the first time when the distractor items in the menu were moderately

[1] Note, that the frequency difference in the number of items visited for positions 9 through 16 in Figure 3.5 probably reflects the difference in the position of the target item across trials.

relevant to the goal statement, compared to not relevant or not at all relevant (*M* = 12.08%, *SD* = 13.60%; *M* = 18.75%, *SD* = 23.92%; *M* = 20.83%, *SD* = 19.52%, respectively). There was a significant main effect of distractor relevance on the proportion of first-visit-selections, *F* (2, 38) = 3.43, *p* = 0.043, *MSE* = .02. Tests of within-subjects contrasts revealed a significant linear trend, *F* (1, 19) = 6.45, *p* = 0.02, *MSE* = .02, suggesting that participants were more likely to select an item sooner when the distractors were less relevant to the goal. The interaction between target and distractor relevance was non-significant, *F* (2, 38) = .75, *p* = .48, *MSE* = .03.

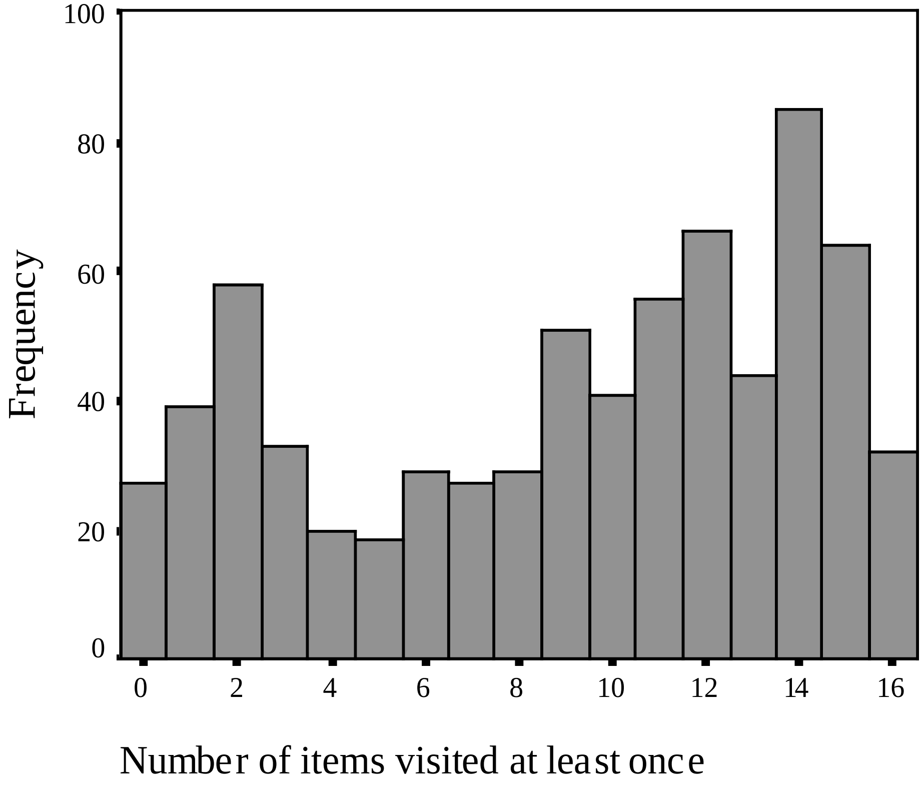


Figure 3.5. Distribution of number of items visited after the initial visit to the selected item.

3.3.2.5. Further strategic adaptations

Earlier analysis of the number of items fixated at least once suggested that people rarely fixated all of the items in the menu prior to selection. In fact, on only 8.19% of searches were all 16 items in the menu visited prior to selection. Obviously, all of the items in the menu were unlikely to have been visited, if an item was selected immediately after an initial visit. Analysis of eye movement protocols, suggest that another reason for participants often not visiting all of the items in the menu, was because they frequently skipped over items as they scanned down to the bottom of the menu. In other words, when an item was fixated, participants often did not fixate the next neighbouring (or spatially contiguous item) in the list, but sometimes jumped to the second item from the current in the list. Analysis explored this observation by considering the probability that a gaze transition occurred between non-contiguous items (e.g., item 3 to item 5). On average 30% (SD = 27.66%) of downward gaze transitions occurred between items that were not spatially contiguous.

I observed that when participants chose not to select the target item after visiting it for the first time, they continued assessing items in the menu, but were more likely to skip over some of the intermediate items as they scanned down to the bottom of the menu. In other words, participants appeared to be more likely to skip items after visiting an item that was highly relevant to the goal statement. To explore this observation a 2 x 2 x 3 (fixation-of-selected-items x target-relevance x distractor-relevance) repeated-measure ANOVA was conducted were an additional factor was included which considered gaze transitions that occurred before and after the initial fixation of the selected item. There was not a significant main effect of whether the selected item had been fixated on the proportion of gaze transitions between spatially non-contiguous items, *F* (1, 19) = 3.73, *p* = 0.07, *MSE* = .03. There was a significant main effect of target relevance, *F (*1, 19) = 5.92, *p* = 0.03, *MSE* = .01. The interaction between target relevance and whether the selected item had been fixated was also significant, *F* (1, 19) = 4.59, *p* = .05, *MSE* = .02.

The interaction is shown in Figure 3.6 that presents the proportion of gaze transitions between spatially non-contiguous items before and after the selected item was first fixated. Further analysis of the simple main effects found that when the target item was highly relevant then there was a significant affect of whether the selected item had been fixated on the proportion of skipping gaze transitions, *F* (1, 19) = 6.93, *p* = .02. Whereas, when the target item was moderately relevant there was not a significant effect of whether the selected item had been fixated, *F* (1, 19) = .12, *p* = .73. There was no significant effect of distractor relevance on the proportion of gaze transitions between spatially non-contiguous items, *F* (1, 19) = 1.81, *p* = 0.18, *MSE* = .03. All other second-order and third-order interactions were also non-significant (for brevity these are not reported).

Finally, I also observed that after first visiting the eventually selected item, participants would sometimes leave the mouse hovering over the item while they scanned over the remaining items in the menu. Interestingly, participants would then select the item with the mouse without first revisiting it (i.e., suggesting that the mouse was strategically left over the item to potentially minimize selection time, if no other competing item were found). This behaviour occurred on approximately 16% of searches.

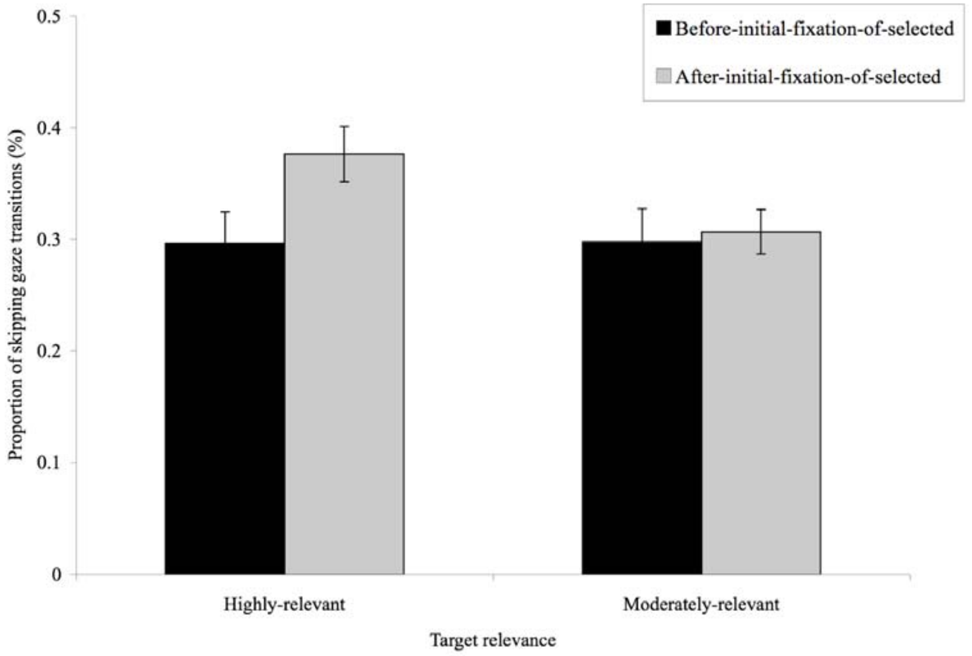


Figure 3.6. The proportion of gaze transitions that were between spatially non-contiguous items before and after the selected item was initially fixated.

3.3.3. Discussion

The results of Experiment 2 show that participants rarely visited all of the items in the available choice set prior to the selection of an item. On only 8.19% of searches were all 16 items in the menu visited prior to selection. This finding is contrary to assumptions implicit in assess-all accounts of web search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2003).

The results of Experiment 2 found that the relevance of all of the assessed items, not just the relevance of the target item, affected both the number of items that were visited, and also subsequently revisited, prior to selection. In particular, participants visited fewer of the available items in the menu when the distractors were less relevant to the goal statement. Further analysis found that this was partially because participants were more likely to select an item immediately after visiting it for the first time (i.e., make a first-visit-selection). This finding suggests that people may make implicit assumptions about the value of items that they *have not assessed* on the basis of generalization from those that they have assessed. Moreover, these results are consistent with context-sensitive accounts (e.g., Cox & Young, 2004; Young, 1998), which assume that people adjust an independent assessment of the relevance of an item in order to derive an estimate that is dependent on the relevance of the other items in the choice set. In contrast, these findings are contrary to threshold accounts (e.g., Miller & Remington, 2004), which predict that participants would make fewer item-visits when the distractors were of greater relevance because the probability that a distractor will exceed the threshold will be increased.

In the current study, participants *frequently* revisited items prior to selection. As already discussed, participants were more likely to revisit items in the menu when the distractors were more, rather than less, relevant to the goal description. Although it may be inevitable that the number of items revisited would be less than the number of items visited at least once, the finding that items that were more relevant to the goal were more likely to be revisited is also consistent with the idea that participants were exhibiting an iterative deepening of attention during search (Rieman, 1994; Rieman, Young, & Howes, 1996; Young, 1998). The results also suggest that revisits to an item were on average for a shorter duration than earlier visits which does not support the idea that participants demonstrated an iterative deepening of attention, involving the use of increasingly costly assessment methods on a few candidate items for selection (Franzke, 1994, 1995; Rieman, 1994; Rieman, Young, & Howes, 1996; Young, 1998). Recall that Rieman, Young, and Howes assume that early visits involve a low quality, low cost assessment of the item, and that subsequent visits involve progressively higher quality, more costly assessment of the item. Perhaps our finding appears inconsistent with the empirical observations of Rieman (1994) and Franzke (1994, 1995) because of the different tasks (i.e., learning to use a computer application vs. searching a web-page for a goal-relevant label) or because of the different experimental methodology (verbal protocol vs. eye-tracking). It is possible that much less elaboration and reification of the meaning of labels is common during web search than it is for people learning a complex computer application package. Either way, there is clearly more to be explained.

In addition to the main findings of the current study, I also observed that participants sometimes skipped spatially contiguous items as they scanned down the list of menu items. Similar skipping behaviour has previously been reported in studies of simple, routine menu selection (Aaltonen, Hyrskykari, & Räihä, 1998; Byrne, Anderson, Douglass, & Matessa, 1999; Hornof, 2004) where participants are required to search a menu for a known target item (i.e., a single number or letter). Interactive search tasks, in contrast, require the participant to estimate the probability that the selection of an option would lead to the goal based on the semantic match of the items label to the goal statement (Blackmon et al., 2002, 2003; Chi et al., 2003, 2001, 2000; Turney, 2001; Farahat, Pirolli, & Markova, 2004). It was unexpected therefore, that people would skip items during interactive search (i.e., where labels must be assessed semantically). The observation suggests that people might be able to make semantic assessments of items, and not just pattern recognitions, based on non-foviated items. More interestingly, it also suggests that people sometimes choose to back off to lower cost assessment methods when they have found a candidate for selection.

In the current study, human relevance judgements were collected in order to provide a more conservative method to estimate label relevance. During the ratings study, all labels for a particular goal statement were presented at the same time. A criticism of this method (Young, personal communication) is that it may be the case that when participants were rating the relevance of a label, judgments could be influenced by the context provided by the other options available at the time of rating, such that judgments of a labels relevance to a given goal statement might not be completely independent. It is arguable that the best way to gather ratings (Young, personal communication) might be by presenting participants with individual pairs of <*goal statement*> : <*label*> in a randomized order across participants. Such a methodology would have the benefit of perhaps providing an entirely independent estimate of the relevance of each label. This method was not employed, however, due to the practical problem of maintaining participant motivation to provide accurate estimates over the duration of what would be a rather lengthy and tedious task.

3.4. Experiment 3

Experiment 3 manipulated the position of the target item within the choice set so as to further test the hypothesis that choosing whether to select or assess more items is dependent on the set of assessments so far made. In particular, Experiment 3 set out to address the question of whether participants would choose to select the target item immediately (i.e., make a first-visit-selection without further assessment) more or less frequently depending on the item's position.

Models of interactive search that rely on a simple selection threshold (MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004) predict that selection should occur when the target item is encountered; therefore the relative position of the target should not influence selection. Models that rely on an assess-all selection strategy (Pirolli & Card, 1999; Pirolli & Fu, 2004) predict that the position of the target does not influence when selection occurs. Whereas, context-sensitive accounts of interactive search (Cox & Young, 2004; Young, 1998) predict that the position of the target item in the menu will affect the likelihood that it is selected immediately.

Cox and Young (2004) have predicted that participants should be more likely to select an item without further assessment if that item occurs earlier, rather than later, in the choice set. The reason for this prediction is because their model assumes that when a goal-relevant item is encountered very early on during search, it is worth investing further in that item by performing a more costly, but higher quality assessment of the item (which would in turn lead to the selection of the item). Whereas, if a goal-relevant item is encountered later, then a low cost, low benefit assessment of the remaining items in the menu is more worth while.

It is possible that Cox and Young's (2004) prediction may not be a necessary consequence of context sensitive theories. A central assumption of context-sensitive accounts of interactive search is that the decision to select an item depends on the quality of the assessments so far made.[2] From this perspective, people should be more likely to select an item when more of the items in the choice set have been assessed, especially if those items are less relevant to the goal. The reason for this is that if it is assumed that as a participant scans down through the list of menu items from the top to the bottom of the choice set, then there is more chance for the distractors to influence selection when the target item is positioned towards the bottom of the menu. If it is also assumed that the net effect of poor distractors is to increase the perceived value of the target then the value of the target should be higher if it is encountered later in the trial.

In addition to predicting that people would make an immediate selection of the target item more frequently after having experienced more distractors, it was also predicted that participants would shift to a skipping strategy after having visited the target item. This prediction was made partly on the basis of observed skipping behaviour in Experiment 2:

[2] An ACT-R model is presented in Chapter 4 that highlights this theoretical claim

Participants often did not select an item immediately, but rather on first visiting the target item shifted from visiting each item in turn to making skipping visits to the remaining items. There is no account of skipping behaviour in previous models of interactive search (MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Fu, 2003). While Cox and Young (2004; Young, 1998) have not reported any prediction for skipping behaviour, the behaviour is at least consistent with their theory of interactive search. An empirical investigation of the models behaviour (described in detail in Chapter 4) predicted that the model may opt to continue assessing the remaining items in the menu with a low cost, low benefit assessment procedure following the assessment of a highly relevant item.

Furthermore, skipping is of particular interest because its presence alongside contiguous assessment could offer quantitative evidence that people strategically deploy multiple assessment methods during interactive search. Only protocol evidence from people learning an application's command menus is available in the literature (Franzke, 1994, 1995; Rieman, 1994).

3.4.1. Method

3.4.1.1. Participants

Sixteen Cardiff University undergraduate psychology students participated in return for course-related credit. None of the participants had taken part in any of the other experiments reported in this thesis. All participants were native English speakers and had normal uncorrected vision. All participants were experienced in using a World Wide Web browser, and all had been required to use various computer software packages to produce coursework.

3.4.1.2. Design

The experiment was a within-subjects design, with two levels of target position and three levels of distractor relevance (moderately relevant, not relevant, or not at all relevant). The target item was either located towards the top of the menu (positions: 2, 3, 4) or towards the bottom of the menu (positions: 13, 14, 15). Each menu contained a single target item that was rated as highly relevant to the goal (i.e., received a median rating of 1 from participants, see Section 2.1 for more details). The manipulation of the position of the target item was counter-balanced across participants. For a given menu, the location of the target item was counter balanced, such that the target item was located towards the top of the menu for half of the participants and was located towards the bottom of the menu for the other half of the participants. The content of the menus differed only in terms of the relevance of the distractor items: In the moderate distractor condition the mean rating of the items was 3; for the poor distractor condition the mean rating of the items was 4; for the very poor distractor condition the mean rating of the items was 5. The primary focus of this study was on eye-tracking data of participant's eye movements up to and including the first selection of an item.

3.4.1.3. Materials and procedure

As in Experiment 2, participants completed 40 interactive search tasks, which required them to search a simplified web page (or menu) for information relevant to a given goal statement. In the current study, the goal statements were the same as those in Experiment 2, and participant's ratings of sampled web-labels (see Experiment 1) allowed us to generate 16-item menus. The menus contained labels of varying semantic relevance to the goal statement: each menu contained a single target item that was rated as highly relevant to goal statement, and differed only in terms of the relevance of the remaining 15 distractor items. The experimental materials were controlled by a purpose-built Microsoft Visual Basic program presented on a high contrast 19 inch FC Trinitron CRT monitor. The menu items were presented in a vertical list in font 15 Cosmic Sans MS and the approximate distance between each label was three degrees of visual angle. Each trial commenced by the participant first reading the goal statement. When the participant was ready, they selected a search button with the mouse, which then removed the goal statement and displayed the menu on the screen. Participants were instructed to scan through the menu items, commencing their search at the top of the menu, and to select the item that they believed to be most relevant to the goal statement as quickly and accurately as possible. As before, in order to impose a meaningful cost structure to the task participants did not progress to the next trial until they selected the target item (i.e., the correct item). If they selected a distractor then they were presented with the same task again. This procedure was repeated until the target was correctly selected. Eye movement data was recorded using an ASL Pan/Tilt optics eye tracking system, which was sampled at a rate of 50 times per second. Eye movement fixations were determined using the same procedure outlined in Experiment 1.

3.4.2. Results

3.4.2.1. Accuracy

In Experiment 3, participants were less likely to accurately select the target item when the distractor items were moderately relevant to the goal statement (M = 79.17%, SD = 18.93%), compared to when the distractors were not relevant (M = 94.27%, SD = 10.03%) or not at all relevant (M = 96.88%, SD = 6.61%). A 2 x 3 (target position x distractor relevance) repeated-measures ANOVA found a significant main effect of distractor relevance on selection accuracy, F (2, 30) = 20.25, p < .001, MSE = .140. The position of the target item did not have a significant effect on accuracy, F (1, 15) = .92, p = .35, MSE = .015, and the interaction was also non-significant, F (2, 30) = 1.07, p = .36, MSE = .021.

For all subsequent analyses, I only consider trials in which the participant correctly selected the target item on their initial selection. This differs from the analysis in the previous experiments (Experiment 1 and Experiment 2) in which all trials were analyzed regardless of the accuracy of the initial selection. The reason for this was that in the present study the position of the target item was an independent variable, therefore it was important to exclude trials in which items other than the target item were initially selected. Only 10% of trials were excluded from further analysis.

3.4.2.2. Time to selection

The time to the selection of the target was less when it was located towards the top of the menu (*M* = 6.52 s, *SD* = 2.93 s) compared to when it was located towards the bottom of the menu (*M* = 9.43 s, *SD* = 2.49 s). The time to selection of an item was greater when the distractors were moderately relevant (*M* = 9.35 s, *SD* = 4.01 s) compared to not relevant (*M* = 7.15 s, *SD* = 2.13 s) or not at all relevant (*M* = 7.43 s, *SD* = 2.33 s). A 2 x 3 (target position x distractor relevance) repeated-measures ANOVA on the time taken from the start of a trial to the initial selection of the target item, found that there was a significant main effect of target position, *F* (1, 15) = 28.267, *p* < .001, *MSE* = 7.179, and distractor relevance, *F* (2, 30) = 17.82, *p* < .001, *MSE* = 2.57. The interaction was not significant, *F* (2, 30) = .12, *p* = .89, *MSE* = 5.08.

3.4.2.3. Proportion of first-visit-selections

It can be seen in Figure 3.7 that participants were less likely to make a first-visit-selection when the target item was positioned towards the top of the menu (*M* = 42.99%, *SD* = 17.60%) than when it was positioned towards the bottom of the menu (*M* = 46.08%, *SD* = 15.50%). Furthermore, Figure 3.7 shows that participants were more likely to select an item immediately after visiting it for the first time when the previously visited distractor items were less relevant to the goal. When the target was located towards the top of the menu, however, the relevance of the distractor items did not affect the decision of whether to select the target immediately (because only a few distractors would have been assessed).

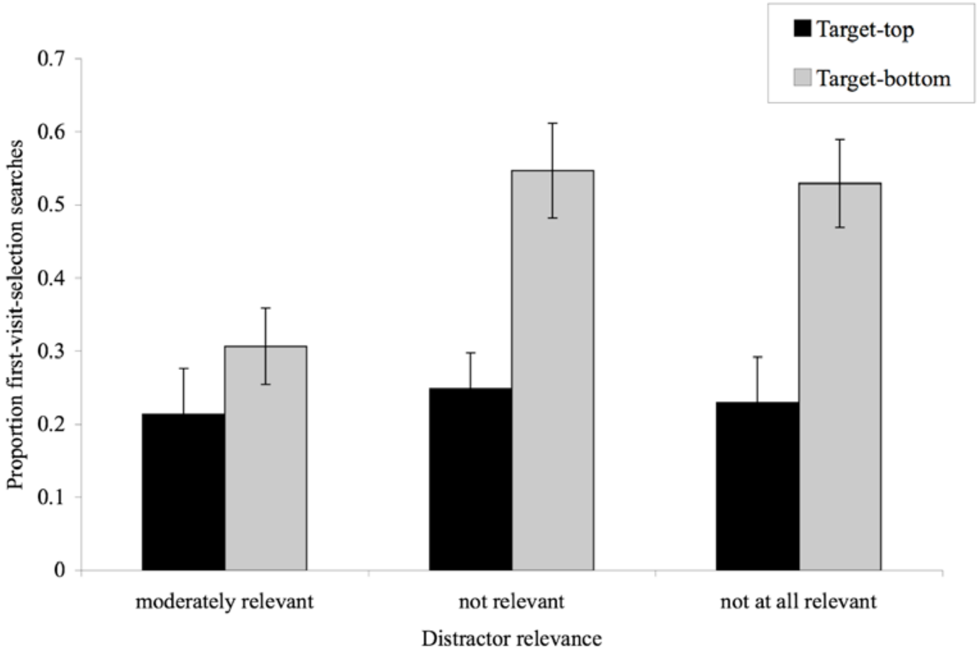


Figure 3.7. The proportion of trials in which the participant selected the target item after visiting it for the first time.

A 2 x 3 (target position x distractor relevance) repeated-measures ANOVA on the likelihood that participant's made a first-visit-selection found a significant main effect of target position, *F* (1, 15) = 18.05, p < .001, *MSE* = .07, and distractor relevance, *F* (2, 30) = 6.76, *p* = .006, *MSE* = .03. The interaction was not significant, *F* (2, 30) = 2.23, *p* = .125, *MSE* = .06, however.

Planned simple effects of distractor relevance were conducted. It was found that when the target item was positioned towards the bottom of the menu participants were less likely to select it immediately when the previously visited distractors were moderately relevant to the goal statement (*M* = 30.62%, *SD* = 20.84%), compared to not relevant (*M* = 54.69%, *SD* = 25.97%) or not at all relevant (*M* = 52.92%, *SD* = 24.25%), *F* (2, 14) = 7.81, *p* = .005. Not surprisingly, when the target was positioned towards the top of the menu the relevance of the distractors did not significantly affect the proportion of first-visit-selections, *F* (2, 14) = .123, *p* = .885.

3.4.2.4. Skipping gaze transitions during interactive search

A skipping gaze transition was defined as any gaze transition that did not occur between spatially contiguous (i.e., neighboring) items. The number of skipping gaze transitions was then divided by the total number of gaze transitions for a given trial. Trials in which the number of gaze transitions were less than or equal to 1 were excluded. Furthermore, in the analysis I only considered gaze transitions that went in a downward direction (i.e., item 2 to item 4). This conservative analysis was adopted because following the analysis of Experiment 2 it was believed that most upward gaze transitions were motivated by the need to verify the location of an item for selection with the mouse, rather than by the need to make new assessments. This amounted to 23.07% (*SD* = 16.43%) of all gaze transitions across participants being excluded because they traveled in an upward direction. It was found that approximately half (*M* = 51.79%, *SD* = 10.52%) of all downward gaze transitions did not occur between spatially contiguous items.

It was predicted that when the target item was located towards the top of the menu, participants would be more likely to skip items during a trial, than when the target item was located towards the bottom of the menu. This prediction was derived from the hypothesis that participants would be less likely to visit every item in turn after visiting, but not necessarily selecting, the target item. It was found that more gaze transitions were between spatially non-contiguous items when the goal was located at the top of the menu (*M* = 54.13%, *SD* = 7.89%) compared to when the target was at the bottom of the menu (*M* = 50.55%, *SD* = 1.74%). A 2 x 3 (target position x distractor relevance) repeated-measures ANOVA found a significant main effect of target position on the proportion of skipping gaze transitions, *F (*1, 15) = 10.86, *p* = .005, *MSE* = .005. The relevance of the distractor items did not have a significant effect on likelihood that participants would skip spatially contiguous items at gaze transitions, *F (*2, 30) = 1.92, *p* = .16, *MSE* = .001, and the interaction was also non-significant, *F* (2, 30) = .08, *p* = .92*, MSE* = .001.

3.4.3. Discussion

Experiment 3 manipulated the position of the target item within the choice set in order to further test the hypothesis that the decision of when to select an item is dependent on the set of assessments so far made. The results of the study suggest that participants were significantly more likely to make a first-visit-selection when the target item was positioned towards the bottom of the menu than when it was positioned towards the top of the menu. Furthermore, when the target was positioned towards the bottom of the menu, participants were more likely to make a first-visit-selection when the previously visited distractors were less relevant to the goal. Not surprisingly, when the goal was positioned towards the top of the menu the relevance of the distractors did not significantly affect the proportion of first-visit-selections. Perhaps, as fewer distractors were visited, their relevance had less affect on the decision of whether to select the target item when it was encountered.

The findings support context-sensitive accounts of interactive search (Brumby & Howes, 2004; Cox & Young, 2004; Young, 1998) that predict that the position of the target item in the menu should affect the likelihood that it is selected immediately. Moreover, the findings of the current study are consistent with Young's (1998) framework for interactive search, which assumes that items are assessed to gain information about which item likely leads to the goal and that this information gain is finite. The idea is that people then terminate search by selecting an item when the expected gain in information in conducting further assessment is not worth the expected cost incurred. If there is a menu which contains a single highly relevant target item then when the target item is located at the bottom of the menu there is less uncertainty as to which item leads to the goal, therefore, there is potentially less information to be gained by conducting further assessments. When the target item is encountered early on during the search process (i.e., prior to the assessment of distractor items) there is less information, therefore, further assessment is likely to lead to a gain in information that will reduce the uncertainty as to which item leads to the goal. Similarly, the presence of distractors that are more relevant to the goal statement has the effect of increasing the uncertainty as to which items should be selected and therefore selection is delayed because further assessment is likely to lead to a gain in information. The findings also support Brumby and Howes' (2004) ACT-R model, which predicts that participants would be more likely to select the target immediately when it is encountered later rather than earlier during the search.

The data do not support Cox and Young's (2004) prediction that participants should be more likely to select an item without further assessment if that item occurs earlier, rather than later, in the choice set. This prediction was made from the assumption that when a goal-relevant item was encountered very early on, the model would invest further in that item by performing a more costly, but higher quality assessment of the item (which would in turn lead to the selection of the item). It seems more likely that when a goal-relevant item is encountered, instead of investing further in that item, people may opt to continue assessing the remaining items in the menu with a low cost, low benefit assessment procedure prior to selection. (This is discussed in more detail below.)

The finding that the relative position of the target item in the menu influenced selection does not support accounts of interactive search (e.g., MacGregor, Lee & Lam, 1986; Miller & Remington, 2004; Pirolli & Fu, 2003). In particular, threshold accounts of interactive predict that selection should occur when the target item is encountered and that the relative position of the target should not influence selection. Similarly the findings of the current study are inconsistent with models that rely on an assess-all selection strategy

The findings of Experiment 3 also provide evidence that participants sometimes choose to skip items during interactive search, and that participants were more likely to skip items when the target was located towards the top of the menu than the bottom of the menu. This finding implies that participants are more likely to skip items after visiting an item that is highly relevant to the goal. In other words, that the strategy is sensitive to the assessment of an item that is a potential candidate for selection.

One interpretation of skipping behavior is that it reflects the use of a low quality, low cost assessment method (i.e., one in which people are rapidly accessing low level information about multiple items within a single eye movement). This interpretation is consistent with Cox and Young's (2004; Young, 1998) accounts of interactive search, and also the related-work of Rieman, Young, and Howes (1996), which assume that people make choices between assessment methods that vary in their costs and potential benefits. This interpretation is also consistent with theoretical assumptions made in various cognitive architectures (e.g., ACT-R: Anderson et al., 2004; Salvucci, 2001; EPIC: Kieras & Meyer, 1997; Hornof, 2004). The details of each approach differ somewhat, however. Hornof (2004) proposes a model in the EPIC cognitive architecture (Kieras & Meyer, 1997) that implements a maximally efficient foveal sweep strategy in which multiple items within the fovea (defined as one degree visual angle) are assessed with a single item visit (fixation). In contrast, Salvucci (2001) has proposed a model, integrated within the ACT-R cognitive architecture (Anderson et al., 2004), in which it is assumed that eye movements are a response to shifts of visual attention. The model prepares and executes an eye movement whenever the eye movement processor becomes available again after the previous eye movement. Attentional shifts take precedence. In Salvucci's model, attention stops on every item, but the eyes only move every two or three shifts of attention, and thus skip over items. It is unclear which of these models provides a better model of the data reported in the current article.

An alternative interpretation of the observed skipping behavior is that participants are only accessing information about the item that is directly foveated and *not* accessing information about multiple items within a single eye movement. It may be the case that participants opt to skip items in order to sample some further items in the choice set in order to evaluate a potential target item against a more accurate estimate of the average relevance of the items in the menu. This view is broadly consistent with Bayesian analyses of interactive search and information foraging (e.g., Pirolli, 2005; Young, 1998). It is unclear why they should sample spatially distributed items, however. Further work is required.

3.5. Experiment 4

The previous three experiments manipulated the relevance of the labels in the immediate (or local) menu choice set. These experiments were motivated to discriminate between different hypotheses concerning strategies for determining when selection of an item should occur during interactive search. Indeed, a key feature of these existing models of interactive search (Cox & Young, 2004; MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2004; Rieman, Young, & Howes, 1996; Young, 1998), with the exception of that reported by Howes, Richardson, and Payne (2002), is that they assume that behaviour is determined entirely by the relevance of the items in the local choice set. Most interactive search tasks clearly occur over repeated search episodes, however. People use the backup button reasonably frequently while searching the web (Catledge & Pitkow, 1995) and there is considerable revisiting of the same web page during browsing (Tauscher & Greenberg, 1997). Experiment 4 was concerned with exploring whether the relevance of the labels in the distant (or previously experienced) choice sets influence subsequent search behaviour.

There is some evidence to suggest that people are sensitive to the relevance of labels that are not present in the immediate choice set but which were experienced in other parts of a web-site (Payne, Richardson, & Howes, 2000; Howes, Richardson, & Payne, 2002). Payne, Richardson and Howes had participants repeatedly locate target items within multi-page web sites that consisted of a series of binary choice menu pages. The study found that when participants revisited a particular choice point to try and find a target item for a second time, navigation choices were sensitive to a variety of information sources. Obviously, participants learned over consecutive trials to recognise that labels were targets and which were distractors. Prior to acquiring such recognition knowledge to discriminate between items, label relevance and familiarity were found to play a key role in guiding navigation choices. In particular, Howes, Payne, and Richardson (2002) found that the decision to leave a page (i.e., to select a backup button) was governed by memory for the quality of the unselected item on the previous menu page.

In response to the empirical data, Howes, Richardson and Payne (2002) proposed a model of interactive search that was sensitive to label relevance outside of the local (or immediate) choice set. The model was of search within a multi-page web site. In the model, the decision of which item to pursue was controlled by a threshold which was sensitive to options that were not necessarily within the immediate choice set of the current menu. The distal search context was provided by an episodic memory of assessments of items that were previously visited during the search episode. The model used a simple utility function to determine whether the value of backing up was greater than the value of an available forward move. As a result, the decision as to whether to leave a page (i.e., to select a backup button) was moderated not only by the relevance of the items on the current page, but also by the relevance of items on the previously visited pages of the current site.

Experiment 4 was designed to further explore the role of non-local influences on interactive search behaviour. In particular, the study was interested in whether the ease with which the target item can be successfully discriminated from distractor items in the distant (or previously experienced) choice sets would affect subsequent search behaviour. Current models of interactive search (Cox & Young, 2004; MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2004; Rieman, Young, & Howes, 1996; Young, 1998) do not speak of the potential role of such distal influences on behaviour.

It is well known in the human problem-solving literature, that choice between decision-making operators is influenced by a combination of information from the current context and the success over past experience (Luchins, 1942; Luchins & Luchins, 1959; Lovett, 1998; Lovett & Anderson, 1996; Reder & Schunn, 1999; Schunn & Reder, 2000). In particular, Lovett and Anderson (1996) had participants repeatedly solve a simple problem-solving task (the building sticks task) that was isomorphic to Luchins' (1942) original water jugs task. The building sticks task required participants to add and subtract lengths of sticks to create a final stick that was equal in length to a target. The task could be approached by an undershoot strategy in which a stick that was shorter than the target had lengths added to it, or an overshoot strategy in which a stick that was longer than the target had lengths subtracted from it. Lovett and Anderson found that over recurring trials participants' choice between these two strategies was influenced by the history of success of using a strategy and the context of the current problem. Importantly, these two sources of information were combined independently: Experiencing failures with a particular strategy lead to a decrease in the selection of that strategy across all future problems. It is an open question whether the choice between operators in interactive search (i.e., assessment vs. selection) is similarly sensitive to successes over past experience.

The aim of experiment 4 was to explore whether the ease with which the target items can be successfully discriminated from distractor items in the distant (or previously experienced) choice sets would affect subsequent search behaviour. If it is, then people should be inclined to select items after fewer gazes if their experience is that selection is more likely to lead to success. Conversely, interactive search behaviour might *only* be sensitive to the relevance of the options in the local menu choice set.

Furthermore, Experiment 4 used a different set of interactive search tasks (i.e., goal statements and menu items) to those used in the previous three experiments. Experiment 4 served as an opportunity to replicate the main experimental finding–that the decision to select an item is sensitive to the estimated relevance of all of the items so far assessed, and not just to the most recent item–with a separate set of language materials (c.f. Clark, 1973).

3.5.1. Method

3.5.1.1. Participants

Thirty-six Cardiff University undergraduate psychology students participated in return for course-related credit. None of the participants had taken part in any of the other experiments reported in this thesis. All participants were native English speakers and had normal uncorrected vision. All participants were experienced in using a World Wide Web browser, and all had been required to use various computer software packages to produce coursework.

3.5.1.2. Design

The experiment was a 2 x 2 (distractor relevance x difficulty of previous trials) mixed design, where the difficulty of previous trials was a between-subjects factor and the relevance of the distractor items in the menu was manipulated as a within-subjects factor. Furthermore, trials were split between filler and critical trials. For filler trials, the discriminability of the target item was manipulated with the aim of inducing differential histories of selection success across participants. For the critical trials, participants searched menus that differed in terms of the semantic relevance of the distractor items. Distractors were rated as either moderately relevant or not at all relevant to the goal statement (receiving a median rating of 3 or 5 from participants, respectively). For all trials, the target item was rated as being highly relevant to the goal (receiving a median rating of 1 from participants). As in the previous experiments, the primary focus was on eye-tracking data of participants' eye movements up to and including their initial selection.

3.5.1.3. Materials and procedure

Experiment 4 used a different set of interactive search tasks (i.e., goal statements and menu items) to those used in the previous three experiments (Experiments 1 – 3) reported in this chapter. More specifically, the web usage survey described in Experiment 1 was repeated in order to derive a separate set of ecologically determined goal statements. The web usage survey was posted to under-graduate students in the School of Psychology at Cardiff University, and from the responses to the survey I was able to derive a set of 34 ecologically derived goal statements. Menu labels for each interactive search task were sampled from various web sites provided by respondents and the semantic relevance of these sampled web-page labels was again determined from human relevance judgments. In the ratings study, 19 Cardiff University undergraduate psychology students judged the degree to which an item was relevant to the achievement of a particular goal statement. Participants took part in the ratings study in return for course-related credit. None of these participants took part in any other part of the experiments reported throughout this thesis. Participants were instructed to estimate the likelihood that selecting a label would lead to the achievement of the goal. Participants made relevance estimates on a 5-point scale, where 1 represented a label that was very relevant to the goal description and 5 represented a label that was not at all relevant to the goal description. To gather the ratings all of the sampled web-labels for a particular goal description were made available at once and participants asked to rate them one-by-one. Participants were instructed to make each relevance estimate *independently* of the estimate of the relevance of the other labels presented on the page.

In the experiment, participants completed 24 interactive search tasks, which required them to search a simplified web page (or menu) for information relevant to a given goal statement. Each menu contained a single target item that was rated as highly relevant to goal statement and the remaining 15 distractor items in the menu varied in their relevance to the goal statement. The experimental materials were controlled by a purpose-built Microsoft Visual Basic program presented on a high contrast 19 inch FC Trinitron CRT monitor. The menu items were presented in a vertical list in font 15 Comic Sans MS and the approximate distance between each label was three degrees of visual angle. Each trial commenced by the participant first reading the goal statement. When the participant was ready, they selected a search button with the mouse, which then removed the goal statement and displayed the menu on the screen. Participants were instructed to scan through the menu items, commencing their search at the top of the menu, and to select the item that they believed to be most relevant to the goal statement as quickly and accurately as possible. As before, in order to impose a meaningful cost structure to the task, participants did not progress to the next trial until they selected the target item. Eye movement data was recording using an ASL Pan/Tilt optics eye tracking system, which was sampled at a rate of 50 times per second. Eye movement fixations were determined using the same procedure outlined in Experiment 1.

In Experiment 4, trials were split between 12 filler and 12 critical trials. Participants either completed filler trials in which the target item was easily discriminated from the distractor items, or filler trials in which the menus contained competing distractor items that decreased the discriminability of the target item. The discriminability of the target item was manipulated by the presence or absence of competing distractor items in the menu choice set (i.e., distractors which were highly relevant to the goal-statement). By varying the discriminability of the target item, the manipulation of the filler trials across participants was intended to differentially affect the success rate of participant's initial selection. All participants completed the same critical trials. For the critical trials, distractor items were either moderately relevant or not at all relevant to the goal statement. The target item on critical trials was located in various menu positions (specifically, positions 1, 2, 3, 6, 7, 8) and was highly relevant to the goal statement. Presentation of filler and critical trials was interleaved, but most of the filler trials occurred during the initial half of the experiment.

3.5.2. Results

The aim of experiment 4 was to explore whether the ease with which the target items can be successfully discriminated from distractor items in the distant (or previously experienced) choice sets would affect subsequent search behaviour. A set of initial analyses therefore considered the effect of manipulating the discriminability of the target item on the filler trials.

3.5.2.1. Accuracy

In the experiment, participants either completed filler trials in which the target item was easily discriminated from the distractor items, or filler trials in which the target item was not easily discriminated from the distractors items (i.e., whether or not the menu contained competing distractor items). As expected, participants were significantly more likely to accurately select the target item on their initial selection if they completed filler trials which were designed to be easy (*M* = 81.48%, *SD* = 12.31%) rather than hard (*M* = 48.15%, *SD* = 13.57%), *t* (34) = 7.72, *p* < .001. No further analyses of the filler trials is presented, as they were merely intended to lead to separate histories of success between the two groups of participants.

For all further statistical analyses of the critical trials from Experiment 4, A 2 x 2 (distractor relevance x difficulty of previous trials) mixed-design ANOVA was employed, where the difficulty of previous trials was a between-subjects factor and the relevance of the distractor items in the menu was a repeated-measures factor. Whether participants had previously completed easy (*M* = 84.72%, *SD* = 17.08%) or hard (*M* = 86.57%, *SD* = 18.18%) filler trials did not significantly affect selection accuracy on the later critical trials, *F* (1, 34) = .53, *p* = .47, *MSE* = .012. Whereas, participants were more likely to select the target item when the distractor items were not very relevant to the goal statement (*M* = 99.07%, *SD* = 3.87%) compared to when the distractor items were moderately relevant (*M* = 72.22%, *SD* = 15.43%). There was a significant main effect of distractor relevant on the accuracy of participants' initial selection, *F* (1, 34) = 93.14, *p* < .001, *MSE* = .006. Excluding trials in which the participant did not accurately select the target item on their initial selection from all further analysis controlled for the effect of distractor relevance on accuracy. The distractor relevance x difficulty of previous trials interaction was non-significant, *F* (1, 34) = .44, *p* = .51, *MSE* = .014.

3.5.2.2. First-visit-selection on critical trials

Figure 3.8 shows the proportion of trials in which participants selected the target after visiting it for the first time. Participants who had previously completed trials in which the target item could be easily discriminated were more likely to select an item immediately (i.e., without visiting any further items) on subsequent critical trials, *F* (1, 34) = 17.58, *p* < .001, *MSE* = .025. Replicating the findings from the previous experiments, it was found that the relevance of the distractor items also affected the proportion of first-visit-selections: participants were significantly more likely to select an item immediately after visiting it for the first time when the distractors are less relevant to the goal statement, *F* (1, 34) = 7.19, *p* = .05, *MSE* = .086. The interaction between distractor relevance and difficulty of previous trials was not significant, *F* (1, 34) = .97, *p* = .33, *MSE* = .025

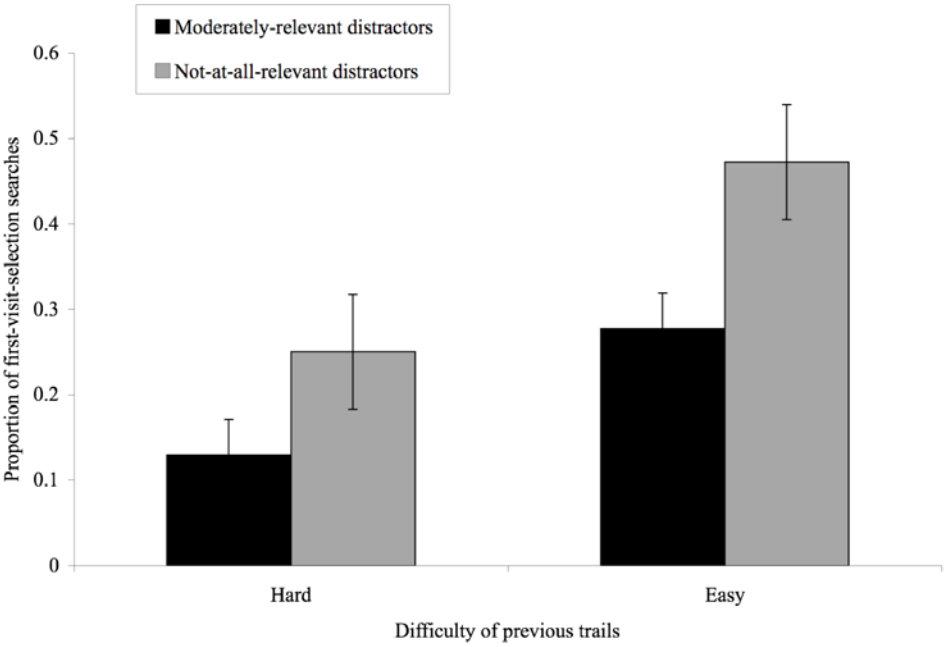


Figure 3.8. The proportion of trials in which the participants selected the target item after visiting it for the first time.

3.5.3. Discussion

Experiment 4 replicated the key finding that the decision to select an item was sensitive to the estimated relevance of all of the items so far assessed, and not just to the most recent item. Experiment 4 used a different set of interactive search tasks (i.e., goal statements and menu items) to those used in the previous three experiments (Experiments 1 – 3) reported in this chapter. Consistent with previous findings, participants were more likely to select an item immediately after visiting it for the first time (i.e., make a first-visit-selection) when distractor items in the menu were less relevant to the goal. In other words, participants visited fewer of the available items in the menu when the distractors were less relevant to the goal statement. This finding is contrary to threshold accounts (e.g., Miller & Remington, 2004), which predict that participants would make fewer item-visits when the distractors were of greater relevance because the probability that a distractor will exceed the threshold will be increased. The finding is consistent with context-sensitive accounts (e.g., Cox & Young, 2004; Young, 1998) which assume that people adjust an independent assessment of the relevance of an item in order to derive an estimate that is dependent on the relevance of the other items in the choice set.

A novel contribution of Experiment 4 is that the results of the study show that the relevance of the labels in the distant (or previously experienced) choice sets influenced participants subsequent search behaviour. In the study, the ease with which the target item could be successfully discriminated from surrounding distractor items was manipulated. Some participants experienced trials in which initial selections were more likely to be correct while others experienced trials in which initial selections were frequently incorrect. The difficulty of these previous trials affected participants search behaviour on subsequent trials (which were the same for both groups of participants). Participants were more likely to select an item immediately after visiting it for the first time when their experience was that selection was more likely to lead to success (i.e., because menu choice sets did not contain competing distractors). Existing models of interactive search (Cox & Young, 2004; MacGregor, Lee, & Lam, 1986; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2004; Rieman, Young, & Howes, 1996; Young, 1998) cannot directly account for the findings of Experiment 4. Cox and Young's and Pirolli and Fu's models can potentially be refined, however.

As outlined previously Cox and Young's (2004; Young, 1998) theory assumes that people make choices between assessment methods that vary in their costs and potential benefits. At the heart Cox and Young's model, are calculations that rely on conditional probability distributions of the potential outcomes of assessments (i.e., the assumed potential benefits of conducting a particular assessment). The model does not currently speak of the process by which these conditional probabilities are acquired. Young (personal communication) assumes that these values arise from long- and short-term experience with the task environment and it is possible that a simple Bayesian up-dating process might capture the calibration of these conditional probabilities. Consequently, the model may be sensitive to the history of experienced relevance. Whether the model would account for the findings of Experiment 4 in an empirically plausible way remains to be seen.

A similar proposal to account for the influence of labels experienced in the distal context on subsequent search behaviour has been outlined in a recent revision to the original SNIF-ACT model (Fu, in press). In the model, called SNIF-ACT 2.0 to reflect the up-grade, a simple utility function is used to determine whether the value of selecting the currently attended item is greater than the expected value of conducting further assessments. The value of selecting the currently attended item is simply the relevance of the label to the goal statement. The *expected* value of conducting further assessments is based on the relevance of the previously assessed items in the local and also previously experience choice sets (i.e., defined as a running aggregate of experienced relevance judgments from past assessments). In this way, the model tends to selects items that were judged to be more relevant to the goal than those typically encountered.

A common feature of the solutions to dealing with the influence of prior experience on subsequent behaviour is for models of interactive search (Cox & Young, 2004; Fu, in press) to be sensitive to the experienced distribution of relevance. It may be the case that participants in Experiment 4 strategically learnt to favour or disfavour a strategy for immediate selecting attractive items. Such an account is consistent with theoretical explanations of how people choose between competing operators during problem solving (e.g., Lovett, 1998; Lovett & Anderson, 1996). Indeed, learning mechanisms that underlie cognitive architectures, such as ACT-R (Anderson et al., 2004; Anderson & Lebiere, 1998) and Soar (Newell, 1990; Rosenbloom, Laird, & Newell, 1993) are sensitive to the successes and failures of basic cognitive operators (i.e., production rules, see modelling section for more details). If it is the case that peoples' interactive search behaviour is also sensitive to the correctness of their selection, then further work is still required in model development.

3.6. General Discussion

In summary, the interactive search experiments reported in this chapter manipulated the relevance and location of items and measured eye fixations up until selection. These measures were used to calculate the number of visits made to each item, and in turn infer which items were assessed. Experiment 1 was an exploratory study, which attempted to use an automated tool (LSA), to determine label relevance in order to set up experimental conditions. Experiment 1 was considered a failed experiment, in the sense that, for a number of reasons discussed earlier, LSA did not provide a reliably accurate measure of relevance. Consequently, human relevance judgements were used throughout the remainder of the empirical studies.

In Experiment 2, it was found that participants were more likely to select an item without visiting any further items in the choice set when the items previously visited were less relevant to the goal. In Experiment 3, participants were more likely to select an item without further visits after more items in the local choice set had already been visited. Experiment 4 found that participants were also more likely to select an item without further visits when their history of previous selections in the experimental session had been more successful (i.e., they had experimented less difficult previous trials). Overall, these findings support context sensitive accounts of interactive search (Cox & Young, 2004; Young, 1998) but do not support accounts in which it is assumed that there is a selection threshold (Miller & Remington, 2004), nor accounts in which thresholds are revised downward if initially set too high (Miller & Remington, 2004), nor accounts in which it is assumed that all items are assessed once prior to selection (Pirolli & Fu, 2003). Moreover, the results of the studies suggest that during interactive search people are sensitive to the ease with which the target items can be successfully discriminated from distractor items in the local (or immediate) choice set, and also in the distant (or previously experienced) choice sets.

Throughout all of the experiments, it was found that participants rarely visited all of the items in the available choice set prior to the selection of an item. This finding suggests that it is unlikely that people adopt a strategy of visiting all of the items in the choice set at least once prior to revisiting the item with the greatest relevance in order to select it (e.g., Pirolli & Fu, 2003). It will sometimes be the case that all items are visited prior to selection but less through strategic design than through an interaction of continual analysis of the utility of selection and assessment with the particular items available on a web page.

In the light of the evidence provided by the experiments presented in Chapter 3 paper it also seems unlikely that people adopt a simple threshold model. According to these models (e.g., Miller & Remington, 2004), when distractors are more relevant people should look at fewer items, because it is more likely that one of the distractors will be above threshold. Data in the experiments indicate that this is not the case, however, if anything people visit more items when the distractors are more relevant to the goal statement.

Participants were found to frequently revisited items prior to selection. The finding that items that were more relevant to the goal were more likely to be revisited partially supports the idea that participants were exhibiting behaviour consistent with the notion of iterative deepening of attention (Franzke, 1994, 1995; Rieman, 1994; Rieman, Young, & Howes, 1996; Young, 1998). The duration of revisits to an item were found to be on average shorter than earlier visits, however. This latter finding does not support the idea that when a goal-relevant item is encountered people invest further in that item by performing a more costly, but higher quality assessment of the item.

While the current studies did not find evidence that people revisit items to assess them with increasingly costly assessment methods they did find evidence that people use more than one kind of assessment procedure. When a goal-relevant item was located, participants were found to sometimes choose to check the remaining items in the menu, but that they were more likely to skip some of these items. In particular, Experiment 3 found that participants were more likely skip items when the target was located towards the top of the menu, compared to when the target was located at the bottom of the menu. Implying perhaps that after encountering a goal-relevant item, participants opted to continue assessing the remaining items in the menu with a low cost, low benefit assessment procedure (i.e., one in which people are rapidly accessing low level information about multiple items within a single eye movement). This finding suggests another way in which interactive search behaviour is sensitive to the relevance of the items so far visited. In this respect the data are consistent with models in which it has been assumed that people have a repertoire of potential assessment procedures available (Cox & Young, 2004; Rieman, Young, & Howes, 1996; Young, 1998). Moreover, considering the semantic similarity between a labelled link and a users information goal (e.g., Blackmon et al., 2002, 2003; Chi et al., 2003, 2001, 2000; Church & Keane, 2004; Kaur, & Hornof, 2005; Pirolli & Fu, 2003; Pirolli & Card, 1999; Miller & Remington, 2004) in and of itself would not appear to be sufficient to explain why people would sometimes choose to skip items and sometimes choose to revisit items.

Previous models of interactive search cannot directly account for the findings of Experiment 4, which demonstrated that the relevance of the labels in the distant (or previously experienced) choice sets influenced participants subsequent search behaviour. In particular, participants were more likely to select an item immediately after visiting it for the first time when they had experienced trials in which initial selections were more likely to be correct. Whereas, participants that had experienced trials in which initial selections were frequently incorrect were subsequently less likely to select an item immediately, instead opting to continuing assessing items in the menu.

Previous model can potentially be refined to account for the affect of the difficult of previous trials on subsequent search behaviour (e.g., Cox & Young, 2004; Fu, in press;

Pirolli & Fu, 2003; Young, 1998). A common feature of the solutions to dealing with the influence of prior experience on subsequent behaviour is for models to be sensitive to the experienced distribution of relevance. It may be the case, however, that participants in Experiment 4 strategically learnt to favour or disfavour a strategy for immediate selecting attractive items. Such an account is consistent with theoretical explanations of how people choose between competing operators during problem solving (e.g., Lovett, 1998; Lovett & Anderson, 1996). Disentangling whether participants were sensitive to the history of successful selection (Lovett & Anderson, 1996) or the history of experienced relevance (Fu, in press) is not a straightforward feat as they were confounded in Experiment 4. Further empirical work is required to separate the influence of these separate sources of prior information on search behaviour.

CHAPTER 4: MODELLING OF INTERACTIVE SEARCH

The aim of Chapter 4 was to explore how and also whether cognitive models of interactive search can account for the main experimental results reported in the previous chapter. As a brief summary, the main findings from Chapter 3 were that: 1. Participants typically did not visit all of the items in the choice set and often revisit items prior to selection, and fewer items were visited when the distractors were less relevant to the goal statement; 2. Participants' were also more likely to select an item immediately when it was located towards the bottom of the menu, particular when the distractors that preceded it were less relevant to the goal; 3. Participants frequently skipped spatially contiguous items as they scanned down the menu, and were more likely to skip items after encountering a highly relevant item. Experiment 4 found that the relevance of the labels in the previously experienced choice sets was found to influence participants subsequent search behaviour. The cognitive models presented in the current chapter do not attempt to deal with this empirical finding, however.

In order to account for these empirical findings a series of models are presented. As a starting point, Cox and Young's (2004) rational analysis of interactive search is described. The current chapter explored whether this rational account might be able to capture the skipping behaviour observed during the empirical studies. Although there were not any theoretical amendments proposed to Cox and Young's model, a novel exploration of the space of the model's behaviour suggests that the skipping behaviour observed during the empirical studies might be a rational adaptation to the task environment. In other words, I ran Cox and Young's model without alterations to their theory. Running the model resulted in a novel observation of its behaviour.

Given the rational account of interactive search (Cox & Young, 2004; Young, 1998), a cognitive model was developed within the ACT-R architecture (Anderson et al., 2004). The model differed significantly from previous ACT-R models of interactive search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2003) by proposing a novel implication of the ACT-R theory of declarative memory for assessment. In the model, the value of selecting an item was sensitive to the context provided by the previously visited items in the choice set.

It is worth commenting that while some previous efforts in modelling have focused heavily on simulating the psychological processes by which relevance judgments are derived (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2004), others (e.g., Miller & Remington, 2004) have been more concerned with the consequences of an assumed distribution of label relevance on search behaviour. This latter approach was favoured in the current context. In particular, the aim was to replicate the key manipulations of label relevance and target position from the experiments in Chapter 3.

4.1. Exploration of Cox and Young's (2004) Model

The experiments presented in Chapter 3 were broadly taken to support context-sensitive accounts of interactive search (Cox & Young, 2004; Young, 1998). It was an open question whether these models can account for the observed skipping behaviour, particularly the finding from Experiment 3 that participants were more likely to skip over adjacent items after assessing one which was highly relevant to the goal. This finding suggests that people may move to a weaker quality assessment of menu labels following an initial assessment of a highly relevant item. While Cox and Young (2004; Young, 1998) have not reported any prediction for skipping behaviour, the behaviour is at least consistent with their theory of interactive search and may be an emergent, but previously unexplored behaviour of the model. This possibility is explored following a detailed description of their model.[3]

4.1.1. Description of Cox and Young's model

Cox and Young's rational framework (Cox & Young, 2004; Young & Cox, 2000; Young, 1998) treats interactive search as an exploratory process in which items in the menu are assessed by performing *exploratory acts* (EAs) in order to reduce the uncertainty as to which option leads to the goal. Each EA is defined by its efficiency (ΔI /C), which is a simple trade-off between the expected information gain ΔI and an estimate of the cost C in time incurred by conducting the EA on an item in the menu. The model chooses to perform at each cycle an EA on an item with the greatest *expected* efficiency (defined shortly). When the expected gain in information from assessment is no longer worth the cost incurred, then the model opts to select the item with the greatest relevance estimate.

As I have already discussed in the introduction, the rational framework assumes that there are a number of task constraints that lead to important theoretical assumptions. First, the model assumes that for each item in the menu ($M_1,\ldots,M_n$) there is an independent relevance estimate R, which reflects a subjective judgment of the likelihood that an item will lead to the goal. Initially, all items in the menu are equally likely to lead to goal (i.e., all $R_i = 1/n$). This assumption implies that initially there is a state of maximum uncertainty as to which item leads to the goal.

Second, the landscape of the relevance values across the items in the menu choice set can be used to define a state of *information*, reflecting a normalization assumption (Young, 1998) in the model. Information is defined as the sum of squared of probability estimates $\sum_{i=1}^{n} P_i^2$ which is calculated by mapping the independent relevance estimates R_i from each item in the menu to a normalized probability estimate P_i, which is defined as $\left(\dfrac{odds(R_i)}{\sum_j odds(R_j)} \right)$. Odds are calculated in the standard way $\left(\dfrac{R}{1-R} \right)$. Basically, if there is only a single item in the menu with a high relevance estimate and all other items have a low relevance estimate, then the sum of squared probability estimates will be high. Whereas, if there are many items with more or less equal relevance estimates, then the sum of squared probability estimates will be low. Moreover, it should be apparent that value of a given menu option is dependent on other the entire set of assessments made.

[3] I am truly grateful to Anna Cox and Richard Young for providing me with a copy of the source code for their model. The model was implemented in Common Lisp.

The model chooses to perform at each cycle an EA on an item with the greatest expected efficiency $\Delta I / C$. Formally, the expected information gain ΔI of performing a given EA is the sum of the expected change in information resulting from each assessment value that can possibly be returned (i.e., $A_1,\ldots,A_n$) multiplied by the probability of the assessment value actually being returned. Each assessment value A_i is defined by the conditional probability that given some assessment of an item it actually leads to the goal p(E+) or does not actually lead to the goal p(E–). For a particular item with a prior relevance estimate R_i, Bayes theorem can be applied to give an updated posterior relevance value,

$$R_i^{'} = \frac{R_i.p(E+)}{R_i.p(E+) + (1 - R_i).p(E-)}$$

Given the posterior relevance estimate $R_i^{'}$ a new set of probability estimates $P_i^{'}$ can be calculated and an expected value of information given $\sum_{i=1}^{n} P_i^{'2}$. The information gain for a particular assessment value applied to a particular item is then $\sum_{i=1}^{n} P_i^{'2} - \sum_{i=1}^{n} P_i^{2}$.

In other words, the model assumes that the expected information gain ΔI of an EA is the average change in information given a small number of discrete assessment values that can potentially be returned by applying a particular EA to an item in the menu.

4.1.2. Model Results

A crucial feature of Cox and Young's model was that different types of assessments (or EAs) could be conducted. Each EA varied in the quality of the assessment that it could potentially return and the expected cost incurred. Young (1998) conceptualizes the quality of an EA in terms of the confidence that a menu item does or does not lead to the goal. In the model higher quality EAs return assessment values that are bimodally distributed (i.e., tending towards relevance values close to 0 and 1) and are associated with greater cost. Lower quality EAs incur less cost, but return assessment values that are normally distributed (i.e., tending towards neutral relevance values, such as 1/N).

For each item the model could choose to perform a *quick-glance*, *semantic*, or *anticipate-selection* EA. The semantic EA represents the process of reading the label and considering the semantic similarity of the labelled link to the information goal and the conditional probabilities were based on participants' ratings (see Experiment 1). The remaining EAs were inspired by the observations of Rieman (1994) that went into the IDXL model (Rieman, Young, & Howes, 1996) and the conditional probabilities were taken from Young (1998). The anticipate-selection EA represents the idea that before people commit to the choice of selecting an item, some additional cognitive activity was involved in trying to anticipate whether the item is likely to lead to the goal. The quick-gaze EA builds on the idea that menu search sometimes involves a simple lexical matching process (Byrne, 2001; Franzke, 1995; Hornof, 2004), more specifically, for the purposes of this discussion it is assumed that a quick-gaze represents a skipping gaze transition. The conditional probabilities and costs for each of these EAs are presented in Table 4.1.

In order to explore whether Cox and Young's (2004) model might be able to capture the skipping behaviour observed during the empirical studies, the model was run on a menu that represented a 16-item menu in which there was a single goal item that was highly relevant to the goal statement and in which the remaining distractor items in the menu were not at all relevant to the goal statement. Table 4.2 presents a run of the model in which the target item was located in the middle of the menu.

It is apparent that after encountering a goal-relevant item, the model opted to continue assessing the remaining items in the menu with a low cost, low benefit assessment procedure. Initially (cycle 1 – 7) the model assessed items in the menu by performing semantic EAs on each item in turn. The assessment of the target item resulted in a large increase in information (cycle 7). At this point, instead of continuing to the next item, the model explored the target item by investing in a more costly, higher quality assessment (by applying the anticipate-selection EA). Surprisingly, the model did not terminate search at this point, even though a highly relevant item had clearly been identified. Instead, the model continued to assess the remaining unassessed items in the menu (cycle 9 – 17), but fell-back on a low cost, low benefit assessment procedure (by applying the quick-glance EA). Search terminated when all the items in the menu were assessed (cycle 17) because the expected information gained by reassessing items is not worth the cost incurred. The model selects the target item because it is the item with the greatest relevance estimate.

Table 4.1. Possible values for each exploratory act (EA) and their conditional probabilities, and cost incurred for model simulation

Quick-glance			Semantic			Anticipate-selection		
	P+	P–		P+	P–		P+	P–
Match	0.40	0.25	Very-relevant	0.79	0.04	Yes	0.90	0.01
Mismatch	0.40	0.80	Relevant	0.18	0.12	Unsure	0.15	0.15
			Moderate	0.03	0.14	No	0.05	0.80
			Poor	0.00	0.18			
			Very-poor	0.00	0.52			
Cost	1.00		Cost	5.00		Cost	10.00	

Note. The columns marked P+ show the probabilities that given the assessment value returned by the EA, the item being assessed leads to the goal and P– if the option does not lead to the goal.

Table 4.2. Model run trace

Cycle	Cost	Menu-item	Assessment	Judgement	Efficiency	R'	Info
0						0.0625	0.0625
1	5	Dist_1	Semantic	Very-poor	0.0054	0.0000	0.0667
2	10	Dist_2	Semantic	Very-poor	0.0057	0.0000	0.0714
3	15	Dist_3	Semantic	Very-poor	0.0061	0.0000	0.0769
4	20	Dist_4	Semantic	Very-poor	0.0066	0.0000	0.0833
5	25	Dist_5	Semantic	Very-poor	0.0071	0.0000	0.0909
6	30	Dist_6	Semantic	Very-poor	0.0077	0.0000	0.1000
7	35	Target	Semantic	Very-relevant	0.0083	0.5683	0.4828
8	45	Target	Anticipate	YES	0.0119	0.9916	0.9900
9	46	Dist_7	Quick-gaze	Mismatch	0.0003	0.0323	0.9905
10	47	Dist_8	Quick-gaze	Mismatch	0.0003	0.0323	0.9911
11	48	Dist_9	Quick-gaze	Mismatch	0.0003	0.0323	0.9916
12	49	Dist_10	Quick-gaze	Mismatch	0.0003	0.0323	0.9922
13	50	Dist_11	Quick-gaze	Mismatch	0.0003	0.0323	0.9927
14	51	Dist_12	Quick-gaze	Mismatch	0.0003	0.0323	0.9933
15	52	Dist_13	Quick-gaze	Mismatch	0.0003	0.0323	0.9938
16	53	Dist_14	Quick-gaze	Mismatch	0.0003	0.0323	0.9944
17	54	Dist_15	Quick-gaze	Mismatch	0.0003	0.0323	0.9950

4.1.3. Discussion

A novel exploration of the behaviour of Cox and Young's (2004) model of interactive search revealed that after encountering a goal-relevant item, the model opted to continue assessing the remaining items in the menu with a low cost, low benefit assessment procedure. This behaviour is consistent with the results of Experiment 3. The model showed that this shift in how people choose to assess items in the menu might be a rational adaptation to the task environment. Moreover, the correspondence between model and data supports the assumption that people might have a repertoire of potential assessment procedures available during the assessment of an item (e.g., Cox & Young, 2004; Rieman, Young, & Howes, 1996; Young, 1998).

4.2. An ACT-R Model of Interactive Search

Cox and Young's (2004) model, described above, relies on a *normalization* assumption (Young, 1998), in order for the subjective value of selecting an item to be sensitive to the context provided by the previously visited item in the choice set. The model provides a rational account (e.g., Anderson, 1990) of behaviour in interactive search in terms of the goals of the cognitive system in relation to the environment.

Following Cox and Young's (2004) rational account of interactive search, a cognitive model was developed within the ACT-R architecture. The benefit of developing a model within a cognitive architecture, such as ACT-R, is that it offers a complimentary approach to rational accounts (see Anderson, 1990, pp. 31) because the models are constrained by the theoretical assumptions of the human cognitive architecture (e.g., ACT-R: Anderson et al., 2004; Anderson & Lebiere, 1998; Soar: Newell, 1990; Rosenbloom, Laird, & Newell, 1993). In order to provide a context in which to understand the key features and processes of the model a brief overview the basic assumptions of the ACT-R cognitive architecture are first presented. It is worth pointing out that the model described here was actually implemented in ACT-R 5.0 (see Anderson et al., 2004, available at: http://act-r.psy.cmu.edu/software/).

The ACT-R cognitive architecture (Anderson et al., 2004) assumes that a central production system coordinates the behaviour of a set of independent modules, each of which are dedicated to processing different kinds of information. For instance, in the current model of interactive search, a visual module might attend to a labelled link, which might trigger the retrieval of a known fact from a declarative memory module. In ACT-R, communication between these independent modules is achieved by passing information through a series of buffers. A goal buffer serves to keep track of the current state-of-affairs insuring that goal-directed behaviour is maintained by guiding the firing of production rules. Production rules are condition-action pairs (i.e., if-then rules) where an action is initiated if and only if the conditions, containing one or more tests of the content of the buffers, are satisfied. At each step only a single production rule can fire. When the conditions of more than one production rule are satisfied, a conflict resolution mechanism based on a simple utility function, derived from Anderson's rational analysis of choice (Anderson, 1990), is used to decide which production rule should be executed.

As will become apparent, an important feature of the current model was that it assumed that the ACT-R goal chunk contained *n* slots, one for each labelled link and each of which, initially, had a value of *unassessed*. These are referred to as *assessment* slots. The goal also had two additional slots, one for the current goal statement and the other for the currently attended visual location. Below is an example ACT-R goal chunk which represents that the task is to do an interactive search task where the goal is to check the expected weekend weather. Let us assume in this example that a visual-object chunk has already been retrieved providing the goal chunk with information about the currently attended item located on the screen by the visual buffer (e.g., the label might be *"Outlook"*).

```
Goal
      ISA          Interactive-search-task
      goal         Check-expected-weekend-weather
      loc          =Currently-attended-visual-location
      menu-pos     =position-in-menu
      current      "Outlook"
      item-1       Unassessed
      ...
      Item-n       Unassessed
      State        "attend"
```

(NOTE: In all examples, the symbol "=" preceding a slot value indicates that it is a variable. Slot values represented as strings do not play an active role in determining chunk activation.)

In the model, the assessment of the currently attended item was achieved by repeated attempts to retrieve chunks from declarative memory. This approach was consistent with that employed in previous ACT-R models of interactive search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2003), but makes a novel use of the mechanism by which source-activation models the focus of attention. Declarative knowledge refers to known facts. In ACT-R, facts are represented as chunks in the declarative module that are defined by a set of feature/value pairs. Anderson and Lebiere (1998) claim that chunks encode "small, independent patterns of information"; more precisely the value of a feature may be a primitive symbol or the identifier of another chunk (where the feature/value pair represent the relation). For instance, a declarative memory chunk representing the knowledge that the word "outlook" is a synonym of forecast and means to predict future weather conditions might be,

```
Outlook
      ISA          semantic-assessment
      text         "Outlook"
      means        Weather-forecast
```

In the model, a series of simple production rules attempted to retrieve various assessment chunks, like the one described above, which associates the currently attended item (`Outlook`) with the goal statement (`Check-expected-weekend-weather)`. An example set of these production rules are described below as both a verbal description of the main features of the rule and also a complete formal specification in the ACT-R syntax.

Assess-relevance

IF the goal is to do an interactive search task to achieve some goal statement and there is visual-object that has been retrieved
THEN retrieve facts about that word

Retrieval-success-assess-further
IF the goal is to do an interactive search task to achieve some goal statement and a fact about that word is retrieved
THEN request retrieval of further semantic-assessment chunks for the currently attended item

Retrieval-failure-item-not-relevant
IF the goal is to do an interactive search task to achieve some goal statement and a fact about that word is not retrieved
THEN remove the current item from the potential choice set and mark as "not-relevant" and find another item to assess.

```
 (P assess-item
    =goal>
      ISA         Interactive-search-task
      goal        Check-expected-weekend-weather
      loc         =Currently-attended-visual-location
      menu-pos    =position-in-menu
      current     nil
      item-1      Unassessed
      ...
      item-n      Unassessed
      State       "attending"
    =visual>
      ISA         visual-object
      Screen-pos  =Currently-attended-visual-location
      Value       =text
==>
    +retrieval>
      ISA         lexical-assessment
      Text        =text
      Means       =word-meaning
    =goal>
      current     =text
      state       "retrieve")

(P Retrieval-success-assess-further
   =goal>
      ISA         Interactive-search-task
      goal        Check-expected-weekend-weather
      loc         =Currently-attended-visual-location
      menu-pos    =position-in-menu
      current     nil
      item-1      Unassessed
      ...
      item-n      Unassessed
      State       "retrieve"
    =retrieval>
      ISA         lexical-assessment
      Text        =text
      Means       =word-meaning
==>
    +retrieval>
```

```
      ISA         semantic-assessment
      word        =retrieval
    =goal>
      item1       =retrieval
      state       “do-further-assessment”)

(P Retrieval-failure-item-not-relevant
    =goal>
      ISA         Interactive-search-task
      goal        Check-expected-weekend-weather
      loc         =Currently-attended-visual-location
      menu-pos    =position-in-menu
      current     nil
      item-1      Unassessed
      ...
      item-n      Unassessed
      State       “retrieve”
    =retrieval>
      ISA         ERROR
==>
    =goal>
      current     nil
      item1       “not-relevant”
      state       “find-next-item”)
```

A crucial component of the model was whether or not a given chunk for the currently attended item was retrieved from the declarative module (i.e., in the above example whether the retrieval module returned a semantic-assessment chunk or an error chunk). In ACT-R, the probability of a chunk being retrieved is depended on its activation. The activation of a chunk is a dynamic value that reflects how often and how recently the chunk was used in the past and how relevant it is to the current context. This definition was vital in order to capture the assumption that the subjective value of selecting an item was sensitive to the context provided by the previously visited item in the choice set.

Formally, the activation *A* of chunk *i* is determined by a combination of base-level activation B_i, spreading activation $\sum_j W_j S_{ji}$, and a transient noise ε .

$$A_i = B_i + \sum_j W_j \times S_{ji} + \varepsilon$$

Base-level activation B_i reflects how often and how recently the chunk was used in the past. This played no functional role in the model and was therefore set to zero. Spreading activation reflects how relevant the chunk is to the current context and played a vital role in determining the models behaviour and is therefore unpacked in more detail.

The amount of spreading activation $\sum_j W_j S_{ji}$ received by a chunk reflects a summation of the attentional weighting W_j of the elements that are part of the current goal, and the strengths of semantic association S_{ji} between the goal statement *j* and chunk *i*. The attentional weights were defined as $W_j = \frac{1}{n}$, where *n* was the number of sources of activation (i.e., the number of sources of activation was dependent on the content of the goal chunk). More precisely, the goal chunk contained slots for $\text{item}_1 \ldots \text{item}_n$ in the menu and the value of these slots could be: *Unassessed*, *Chunk*$_i$, or *"not relevant"*. The initial value of these slots was *Unassessed*. The importance of these different values will be discussed shortly.

The amount of activation received by a chunk also depended on the strength of semantic association S_{ji} between the goal statement *j* and chunk *i*. Following Budiu and Anderson (2004) S_{ji} reflects the similarity between chunk *i* and *j* and was defined as

$$S_{ij} = C + M \times \sigma(j,i)$$

Where, $\sigma(j,i)$ reflects the input measure of semantic similarity between chunks *j* and *i*, and C was a negative quantity that serves as a base of associative strength and M was a positive multiplier. This definition of the similarity between two chunks reflects a simple linear mapping of similarity that varies between 0 and 1, where values closer to 1 reflect greater similarity.

A chunk was only retrieved from the declarative memory module if its activation was greater than a fixed retrieval threshold τ . The activation of a chunk fluctuates with some transient noise ε , which is sampled from a logistic distribution with variance σ. As a consequence the probability *P* that a chunk *j* will be retrieved was defined as $P_j = \frac{e^{A_j/t}}{\sum_i e^{A_i/t}}$, where *t* was a constant dependent on the variance σ of this noise, such that $t = \sigma\sqrt{6/}\pi$. The activation A_j of chunk j also influences the time *T* it takes to retrieve the chunk, such that $T_j = F_e{}^{-A_j}$, where *F* is a constant latency factor; basically, the higher the activation of a chunk the more rapidly it is retrieved from the declarative buffer.

A crucial feature of the model was that different chunk-types were used to represent different types of assessment. The successful retrieval of a chunk was assumed to indicate that there was positive information linking the label and the goal. In the model, this was represented by a change in the value of an assessment slot on the goal chunk. More precisely the initial value *Unassessed* was replaced with the retrieved value *Chunk*$_i$. If during the assessment of an item all of the chunk-types related to that item were successfully retrieved, the model evaluated the item to be highly relevant to the goal statement and it was selected. Clearly, in the case of the example above the label "*outlook*" was likely to be evaluated as highly relevant to the goal statement and would therefore be selected.

Moreover, in the model the value of selecting an item was sensitive to the context provided by the previously visited item in the choice set. This was because the probability *P* of retrieving an assessment chunk from the declarative module was in part dependent on the attentional focus provided by the context of previously assessed items in the menu choice set. The more items that were in the menu choice set the lower the amount of source activation W_j received by a chunk, whereas, the fewer the number of items in the choice set the greater the amount of source activation W_j received by a chunk. Given the current context of the goal some of the chunks associated with an item were retrieved, while others for that same item would fail to be retrieved.

When there was a failure to retrieve an assessment chunk for the currently attended item this resulted in the replacement of a slot value with the string *"not relevant"* (i.e., equivalent to setting the slot value to nil). In other words, this indicated that the item was judged not relevant to the current goal statement and resulted in the removal of a source of activation from the goal chunk (i.e., n was revised to $n - 1$).

In the model items could be reassessed. It should be evident at this stage that even if an item was initially deemed to be not relevant to the goal statement, reassessment of the item may lead to a different evaluation of the items relevance in the future. In particular, given some change in the value of the attentional weight W the probability of retrieving all of the chunk-types related to an item may increase. Consequently, the probability that an item was selected was dependent on all of the items so far assessed, and not just to the most recent item.

The model chose to assess or reassess items in the menu when there was a failure to retrieve an assessment chunk for the currently attended item. In deciding which item to assess next, production rules for assessing the item nearest the current visual location or reassessing some previously assessed item, competed in a stochastic selection process. The utility of the production rule for assessing the item nearest the current visual location was greater than the utility of the production for reassessing an item. In this way some skipping of spatially adjacent items occurred as items were reassessed.

4.2.1. Model Results

The model was evaluated by comparing its behaviour to the results of Experiment 2 and Experiment 3 in the previous chapter. The primary focus of the experiments was on the analyses of participants' eye movement protocols. In evaluating the model, the assumption was made that ACT-R's movements of visual attention to items in the menu can be taken to broadly match participants' eye movement fixations centred on a menu item. To bolster this assumption the model used Salvucci's (2001) model EMMA as a plug-in to the standard ACT-R visual buffer. EMMA provides a more detailed theory of visual encoding by explicitly mapping eye movements to movement of visual attention. Consequently, when the model assessed an item the location of the item in the menu was outputted, therefore, the model's behaviour was subject to the same analysis as participants' eye movement protocols.

The ACT-R model interacted with a menu that was the same as that searched by participants in the experiment. All menus contained 16-items. Each menu contained a single target item and the rest of the items were distractors. In order to model the results from Experiment 2, the position of the target was at first always located towards the top of the list of menu (between menu positions 3 through 8). In experiment 3, the position of the target was manipulated. In order to evaluate the consequences of varying target position, the model was also run on menus were the target was either located towards the top or towards the bottom of the menu. For each trial, the model was run until an item was selected. This differs from the experimental procedure, in which participants did not complete a trial until they had correctly selected the target item. This difference in procedure would not have affected the models behaviour because production rule learning was disabled (such that adjustment were not made to the utility of production rules over consecutive runs). Moreover, the current model was not aimed at addressing any trial-to-trial learning that might occur (i.e., the results of Experiment 4 were not modelled).

4.2.1.1. Model fits with Experiment 2: Effect of label relevance

The model was run on simulated menus which systematically manipulating the relevance of the target and distractor items and compared to the main results from Experiment 2. The model was run through a simplified design that had two levels of target relevance (highly relevant vs. moderately relevant) and two levels of distractor relevance (moderately relevant vs. not relevant). This differs somewhat from the design of Experiment 2, which had three levels of distractor relevance (moderately relevant, not relevant, or not at all relevant). This difference is design was because participants search behaviour was not significantly affected by whether distractors were not relevant or not at all relevant. In comparing the human data to the model, data from the two low relevance distractor conditions was aggregated.

Table 4.3. Estimated Similarity Values.

Label Quality	Chunk-Type		
	Word	Semantic-Assessment	Value-of-Item
Highly-relevant Target	.80	.45	.25
Moderately-relevant Target	.50	.20	.18
Moderate Distractor	.25	.15	.10
Not-relevant Distractor	.20	.10	.05

The representation of label relevance was a major free parameter in the model. The relevance of a label was set by the strength of semantic association S_{ji} between the goal statement *j* and an assessment chunk *i*. Recall that a crucial component of the model was that different chunk-types were used to represent different types of assessment. In the model there were three different assessment chunk-types, therefore, providing 9 free parameters for input similarity values $\sigma(j,i)$ to be estimated. The final input similarity values are presented in Table 4.3. It is worth noting that the input similarity for each chunk-type differed, reflecting differences in the quality of each assessment method. Moreover, these assumptions were consistent with previous models of interactive search (Cox & Young, 2004; Rieman, Young, & Howes, 1996; Young, 1998, see also Table 4.1).

In order to avoid the concern of overly fitting the model to the data, a principled approach was taken to estimating the input similarity values presented in Table 4.3. In order to estimate the input similarity values $\sigma(j,i)$ for each label of the same relevance, values were iteratively estimated for three of the four experimental conditions. For each level of label relevance an input similarity value was obtained that maximized the models fit with the human data across a range of dependent variables over 100 model runs. Once an input similarity value was estimated, I moved to the next condition and obtained a best fitting model by varying input similarity values for the chunk-types associated with only a single label quality. For example, taking the *highly relevant target x not relevant distractor* condition, input similarity values for the chunks associated with highly relevant targets and not relevant distractors were estimated which provided a good fit with the data for that experimental condition. Next, fitting the model to the *moderately relevant target x not relevant distractor* condition, I held constant the estimated input similarity values for the chunks associated with the distractor items and only varied the value of the chunks associated with the target item. In this manner, the models performance on the final condition (*highly relevant target x moderately relevant distractor*) was predictive, at least in the sense that the model was not iteratively fitted to the data.

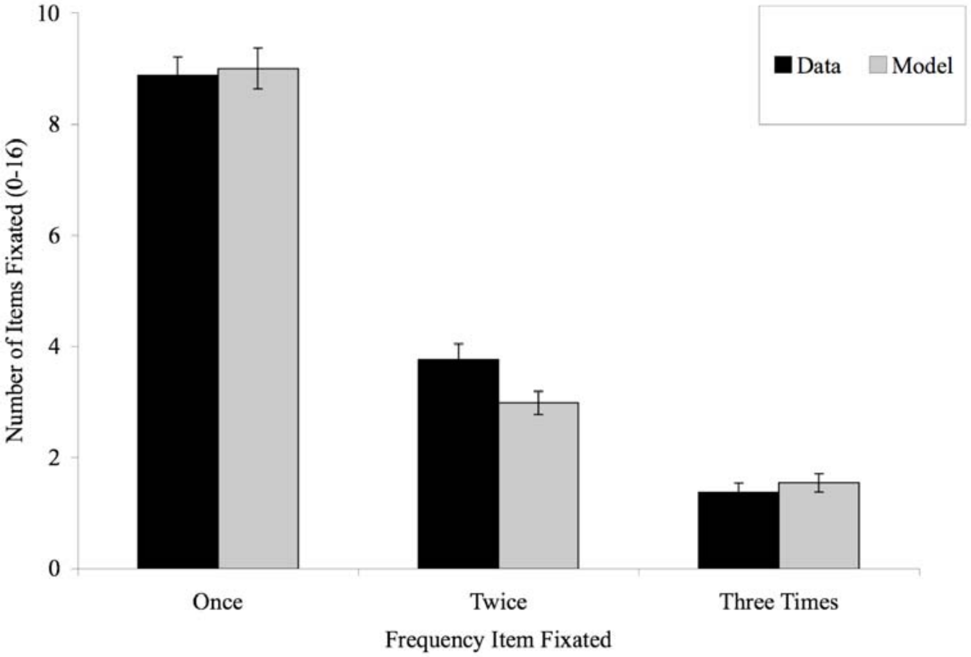


Figure 4.1. Data and model fits for the number of items fixated at least once, twice, and three times. Data has been collapsed across experimental conditions.

Figure 4.1 shows for human data and the model the number of items that were visited at least once, twice and three or more times (aggregated across all experimental conditions). Both data sets show a similar pattern of results: rarely were all of the items in the menu choice set visited prior to selection, and items were frequently revisited on multiple passes prior to selection. This qualitative pattern was an important component of the observed behaviour in the experiments and is one that is not captured by some of the previous models of interactive search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2003).

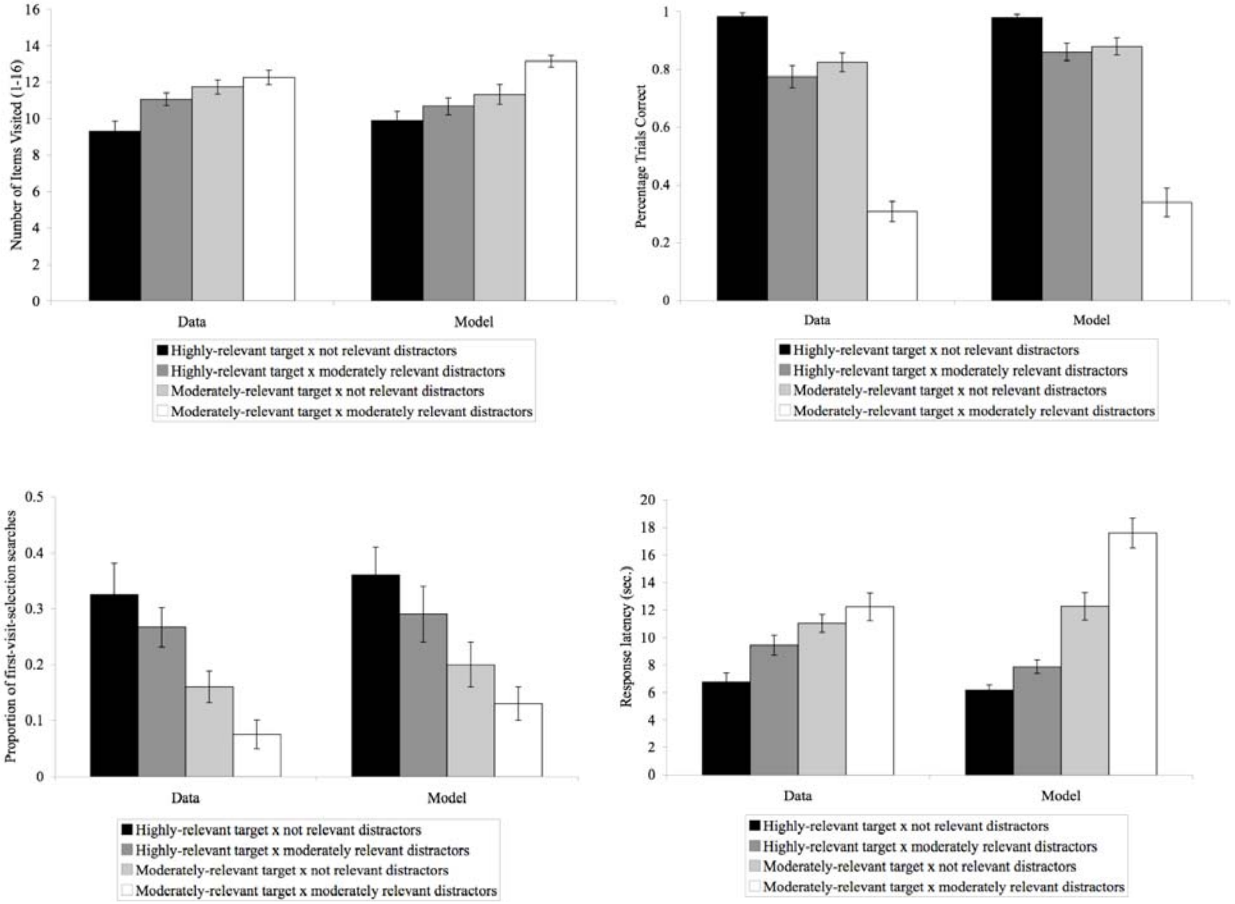
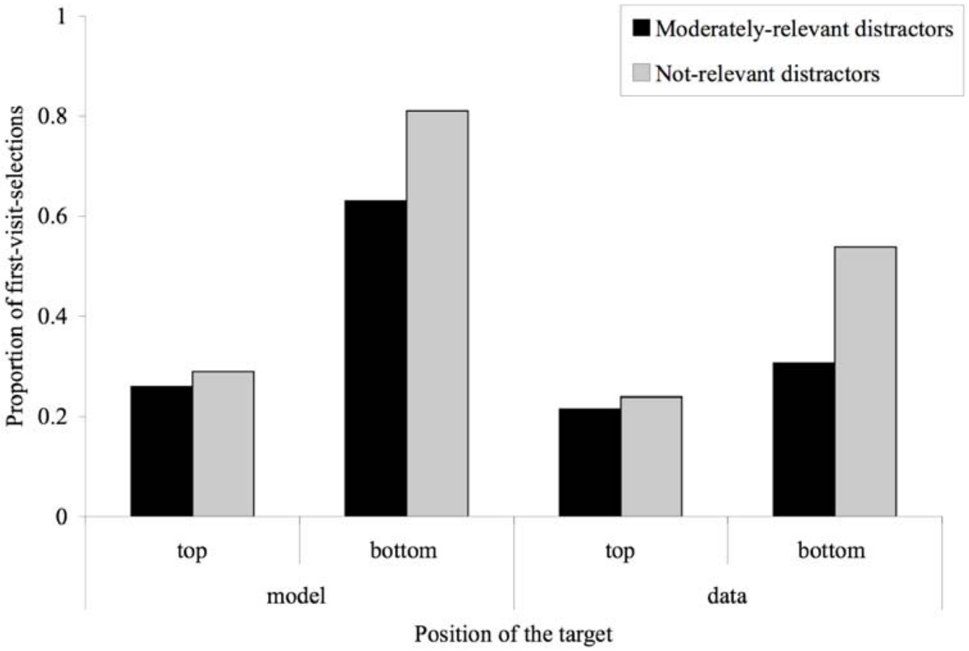


Figure 4.2. Data and model fits for effect of quality of goal and quality of distractor items for the (1) number of items visited at least once; (2) percentage of trials correct; (3) percentage first-visit-selections; and (4) time to selection.

Furthermore, the behaviour of the model was affected by the manipulation of the relevance of the target and distractor items in the menu (or more precisely varying the input similarity values for items, Table 4.3). The model was compared to four dependent variables from Experiment 2. Figure 4.2 shows that the model provided very good fits to the human data for (1) the number of items visited at least once; (2) the percentage of trials correct; (3) the percentage of first-visit-selection trials; and (4) the total time to the initial selection. These quantitative fits were highly significant, $r^2 = .96$, F (1, 10) = 214.08, $p = .001$.



4.2.1.2. Model fits with human data from Experiment 3: Effect of target position

In order to evaluate the consequences of varying target position, the model was also run on menus were the target was either located towards the top of the menu (positions: 3, 4, 5) or towards the bottom of the menu (positions: 12, 13, 14). As in Experiment 3 distractor relevance was also manipulated, but as in the previous section, the model was run through a simplified design that had two levels of distractor relevance (moderately relevant vs. not relevant). None of the models free parameters were varied. In particular, the input similarity values were the same of those estimated in Table 4.3.

Figure 4.3 shows the proportion of trials where the model made a first-visit-selection of an item compared to the human data from Experiment 3. The model was more likely to select the target item immediately when it was positioned towards the bottom of the menu, compared to when it was positioned towards the top of the menu. The model was also especially likely to select the target immediately, when many of the items in the menu had already been evaluated as not relevant to the goal (i.e., when the assessed distractors were poor compared to moderate).

Figure 4.3. Proportion of first-visit-selection trials for data and model.

The experiments presented in the previous chapter found that participants frequently skipped spatially contiguous items as they scanned down the list of menu items. In particular, Experiment 3 found that approximately half (*M* = 51.79%, *SD* = 10.52%) of all downward gaze transitions did not occur between spatially contiguous items. The model did occasionally skip spatially contiguous items in order to reassess items in the menu. These occurred relatively infrequently, with only approximately a fifth (*M* = 18.01%) of all downward gaze transitions occurring between items that were not spatially contiguous.

Furthermore, the results of Experiment 3 found that participants were more likely skip items when the target was located towards the top of the menu, compared to when the target was located at the bottom of the menu. Whether the model chose to reassess items, and, therefore skip over spatially contiguous items', was not sensitive to the position of the target item within the menu.

4.2.2. Discussion

A context-sensitive model of interactive search was developed. The model was implemented in the ACT-R cognitive architecture and considers interactive search from the perspective of attentional focusing. The approach taken differs significantly from previous ACT-R models of interactive search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2004). The model exploited the fact that in the ACT-R cognitive architecture a fixed amount of source activation is distributed among the declarative chunks that are associated with the goal. The probability of retrieving chunks that associated a label with the goal statement was partly dependent on the number of other labels in the choice set which were also relevant to the goal. Consequently, the probability that an item was selected was sensitive to the estimated relevance of all of the items so far assessed, and not just to the most recent item. The models behaviour was therefore consistent with the rational analysis of interactive search provided by Young (1998; Cox & Young, 2004).

The model was evaluated by comparing its behaviour to the human data from the experiments presented in Chapter 3. The model was run on simulated menus that systematically manipulated the relevance of the target and distractor items (consistent with Experiment 2) and also the position of the target item in the menu (consistent with Experiment 3). Overall, the model provided a good fit with the human data. The model rarely visited all of the items in the available choice set prior to the selection and was more likely to select the target item when the surrounding distractor items were less relevant to the goal statement (i.e., when the distractors were poor compared to moderate). In general, the model was more likely to make a positive assessment of an item as more of the available items in the choice set were assessed and consistent with the data the model was more likely to select the target immediately when more it was positioned towards the bottom of the menu compared to towards the top of the menu.

The model was inconsistent with the data in at least one respect. It did not provide a good fit to proportion of skipping downward gaze transitions that occurred between spatially non-contiguous items. The model would skip spatially contiguous items when it chose to reassess a particular item. The data from Experiment 3 found that participants were more likely skip items when the target was located towards the top of the menu, compared to when the target was located at the bottom of the menu. This finding was taken to imply that perhaps after encountering a goal-relevant item people opt to continue assessing the remaining items in the menu with a low cost, low benefit assessment procedure (i.e., one in which people are rapidly accessing low level information about multiple items within a single eye movement). This strategic control in determining how an item was assessed was beyond the scope of the model.

4.3. General Discussion

Chapter 4 presented models of interactive search that accounted for the main experimental findings reported in the previous chapter. Cox and Young's (2004) context-sensitive account provides a rational account of interactive search and were supported by the results of the experiments in Chapter 3. In the current chapter, the model was described in detail and a novel exploration of its behaviour suggested that strategic shifts in how people assess items following the assessment of a goal-relevant item might reflect a rational adaptation to the task environment.

Following the rational account provided by Cox and Young (2004) an ACT-R model (Anderson et al., 2004) of interactive search was developed where the value of selecting an item was sensitive to the context provided by the previously visited item in the choice set. Consistent with the empirical data the model rarely visited all of the items in the available choice set prior to the selection of an item and was more likely to select the target item when the surrounding distractor items were less relevant to the goal statement. The model was also more likely to select the target immediately when it was encountered later rather than earlier during the search process.

The model differed substantially from previous ACT-R models of interactive search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2003) by proposing a novel use of the mechanism by which source-activation models the focus of attention. The model construed assessment as attentional focusing by assuming that activation is distributed between declarative representations of the set of possible selections; the more and stronger the links between the goal statement and the representation of an item in declarative memory the higher that representation's activation. An item was selected once its activation exceeds a threshold. The idea of assessment as attentional focusing may seem counter-intuitive. In the model the goal of assessing unassessed items reduces the probability of retrieving information about the currently attended item. Although, this mechanism predicts the observed behaviour it seems counter-intuitive because, given that the goal is presumably under strategic control, an implication is that participants deliberately reduced the probability of retrieval of information associating an item with the goal (at least initially) in order to achieve the desired overall search strategy.

Although the ACT-R model was influenced by Cox and Young's (2004) rational account, there are important differences between the two models. Cox and Young's (2004) model assumes that assessment of items continues until the cost of performing an assessment outweighs the estimate of the information to be gained from further assessment. In more general terms, the ACT-R model can be thought of as a normalisation + threshold account of interactive search. In the model, the decision to select an item was governed by a dynamic threshold, in which the item that eventually exceeded the threshold gained activation through the assessment of alternative choices. The models make different predictions for how the number of labelled links on a particular web page affects the choice between selection and further assessment. This is explored in more detail in Chapter 5.

CHAPTER 5: IMPLICATIONS FOR DESIGN

One area in which models of interactive search have been applied is to predict the optimal structure to organise the links on a web site so as to reduce users average navigation time to reach the desired target link. Essentially, the content information of a web site can either be distributed over many pages, each containing few labelled links, generating a broad and shallow structure. Alternatively, the same content information can be distributed over fewer pages, each containing many more links, generating a deep and narrow structure. This depth vs. breadth design question has been the subject of much empirical and theoretical attention over a number of years and has resulted in design guidelines that have sometimes been contrary in their recommendations (e.g., Katz & Byrne, 2003; Larson & Czerwinski, 1998; Lee & MacGregor, 1985; Miller & Remington, 2004; Nielsen, 1999; Parush & Yuviler-Gavish, 2003; Shneiderman, 1998).

A number of early empirical studies, which predated the invention of the web, evaluated a variety of hierarchical menu structures in terms of the fastest search times (e.g., Kiger, 1984; Lee & MacGregor, 1985; Norman, 1991; MacGregor, Lee & Lam, 1986; Miller, 1981; Snowberry, Parkinson, & Sisson, 1983). Empirical results of menu search experiments have found that structures with as many as eight selections per page produce faster search results than deeper structures with fewer selections per page (Kiger, 1984; Miller, 1981; Snowberry, Parkinson, & Sisson, 1983) and broader structures with more than eight selections per page produce slower search times (Miller, 1981; Snowberry et al., 1983). The results from a study by MacGregor, Lee, and Lam (1986) were particularly noteworthy because they suggest that people are more inclined to select an item immediately after visiting for the first time as the number of options on a particular menu page increase.

Despite these previous findings from the menu search literature, some have argued that it is good design practice to put more, rather than fewer, links on a web page (Nielsen, 1999; Larson & Czerwinski, 1998). The rationale behind such an assumption seems to be that favouring broad and shallow structure flattens the hierarchy and thereby avoids the cost overheads incurred by users searching and backing up through a deeper site structure. Larson and Czerwinski (1998) examined user search times in web pages of differing hierarchical depth. In contrast to the results from the menu selection studies, they found that users took significantly longer to find items in a three-tiered, 8-links-per-page (8 × 8 × 8) structure than in comparable two-tiered structures with 16 and 32 links per page (16 × 32 and 32 × 16).

Miller and Remington (2004) accounted for the apparent discrepancy between the findings from the menu search literature (Kiger, 1984; Lee & MacGregor, 1985; Norman, 1991; MacGregor, Lee & Lam, 1986; Miller, 1981; Snowberry, Parkinson, & Sisson, 1983) and the result from the web navigation study by Larson and Czerwinski (1998) by showing how performance is affected by an interaction between label relevance and site structure. An engineering model was developed by Miller and Remington which predicted how long it will take a person to find the information that they require given a description of the labels and their connectivity. The aim of the model was to predict the optimal structure of a web site (optimal depth and breadth) for a given level of label relevance. As described previously, for each page items were assessed, and if an item exceeded a threshold it was selected. If when all the items on a page were assessed, none exceeded the threshold, then the threshold was lowered and the items on that page re-evaluated relative to the new, lowered threshold. The model backed up to the previous page in the site when none of the items on a given page exceed the reduced threshold. The model predicted that when the target item was clearly discriminable from the surrounding distractor items, on each page on the path to the goal, a deeper and less broad structure lead to faster search times. When target item was not clearly discriminable from the distractor items, on each page on the path to the goal, a broader and less deep site structure lead to faster search times. An empirical study was reported which corroborated the predictions of the model.

The aim of Chapter 5 was to discriminate between different models-based predictions of the affect of varying the number of labelled links on a page for assessment and selection. Miller and Remington's (2004) model, which predicted the average time to navigate a site structure, assumes that search of each page was based on a threshold strategy. The selection threshold was not sensitive to the number of options on a particular web page. There is at least some empirical evidence from MacGregor, Lee, and Lam's (1986) menu search study, to suggest that the number of options on a particular page affected whether participants chose to select an item immediately (i.e., self-terminating search) or made further assessments prior to selection (i.e., exhaustive search). The results of the experiments presented in Chapter 3 demonstrated that a threshold account fails to capture performance on single page search tasks. Moreover, the empirical findings and the model developed in the previous chapters suggest that varying the number of labelled links on a page might have consequences for the choice between assessment and selection.

5.1. Model-based Predictions

Predictions can be derived from existing cognitive models of interactive search (Cox & Young, 2004; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Young, 1998) of the potential affect on search behaviour of varying the number of items on a particular web page. Assess-all accounts of interactive search (Pirolli & Card, 1999; Pirolli & Fu, 2003) have not been applied to addressing questions of breadth and depth; nonetheless it can be assumed that the models assume that all items are assessed regardless of menu breadth.

Threshold accounts of interactive search, such Miller and Remington's (2004) model, predict that the number of items assessed prior to selection will increase proportionally with the number displayed on the page, assuming that the remaining distractor items never compete for selection with the target (i.e., there is high target discriminability) and the target item is randomly positioned within list of labelled links. Whereas, if the distractor items compete for selection with the target (i.e., there is low target discriminability) then proportionally fewer items should be assessed prior to selection as the number of items increase. It is an open empirical question whether the number of items on a page affects target discriminability.

Context-sensitive accounts of interactive search (e.g., Cox & Young, 2004; Young, 1998) have not hitherto been directly applied to addressing the question of how varying the number of items available on the page might affect search behaviour. Predictions are certainly within the scope of these models and can be readably derived. Cox and Young's model assumes that selection occurs when the expected information gained through further assessment is no longer worth the cost incurred. It should be apparent from the description of the model presented in Chapter 4, that as the number of items on a given web page increase the proportion of information that is gained through the assessment of a particular item decreases, whereas the cost of an assessment remains the same regardless. This is because items rapidly acquire values that are too low to justify assessment; therefore, as the number items on a given web page increase proportionally fewer items should be assessed prior to selection.

In contrast, the ACT-R model described in Chapter 4 predicts that with an increase in the number of items on a page people should prefer to defer the selection of an item, in order to assess more of the available options. Although both models rely on a normalization assumption, which as described by Young (1998), reflects a constraint imposed by the structure of the task environment. The models differ in terms of the assumptions underlying the choice between selection and further assessment. The ACT-R model can be thought of as a normalisation + threshold account of interactive search. In the model the evaluation of a label's relevance was in part dependent on the attentional focus provided by the context of the previously assessed items, therefore, the item that eventually exceeds the selection threshold gained activation through the assessment of these alternative choices. As the number of items in the menu choice set increases, the attentional focus becomes more distributed, therefore, reducing the likelihood that an item would be evaluated as relevant enough to warrant selection.

5.2. Experiment 5

Experiment 5 manipulated the number of items (i.e., breadth) and also the depth of the labels within the web site structure (i.e., the distance from the target) on a single web page. In the current experiment, label relevance was not itself manipulated. To ensure that the distribution of relevance accurately reflected in the current experiment, the content of the web pages used in the experiment was taken from actual web sites. It is an open empirical question whether the distance labels are sampled from the target affects label relevance.

5.2.1. Method

5.2.1.1. Participants

Twelve Cardiff University undergraduate psychology students participated in return for course-related credit. None of the participants had taken part in any of the other experiments reported in this thesis. All participants were native English speakers and had normal uncorrected vision. All participants were experienced in using a World Wide Web browser, and all had been required to use various computer software packages to produce coursework.

5.2.1.2. Materials

The materials were based on real web pages. Ecological validity was important in the design of the current experiment. For this purpose web directories were primarily sampled, such as the DMOZ Open Directory Project (http://dmoz.org), because they provided a hierarchical structure covering a large and broad collection of content links. For each of the sampled pages, the author generated a goal statement. A target item that satisfied the goal statement was identified within the web site. In all, thirty-six different web pages were sampled. To manipulate menu breadth, web pages were sampled which contained approximately the desired number of labelled links. In order to manipulate the distance to the target node, a target item that satisfied a given goal statement within the actual web site was located. For the close condition, the labels from the page in the web site that contained the target were sampled. For the medium, condition the labels from the page in the web site that preceded the page containing the target were sampled. Finally, for the far condition, the labels that were sampled were from a parent page that was at least two-clicks away from the target node in the web site hierarchy. Moreover, no systematic control of label relevance was employed in the current study, for instance to determine whether the target item was indeed relevant to the goal statement.

5.2.1.3. Design

The experiment was a within-subjects design and manipulated the number of labelled links displayed on a page (menu breath was 8-, 16-, or 32-items) and also the depth of the labels in the web site structure across three levels (distance from the target node was close, medium, or far). Each web page contained a single target item that was randomly positioned amongst the distractors on the page. As in the previous experiments, the primary focus was on eye-tracking data of participants' eye movements up to and including their initial selection.

5.2.1.4. Procedure

In the current experiment, participants completed 36 interactive search tasks, which required them to search a simplified web page for information relevant to a given goal statement. The experimental materials were presented as simplified web pages; each page was recreated in html following a standardised format of presentation in which all frames and stylistic features were removed. These simplified web pages were presented using an instance of an Internet Explorer browser presented on a high contrast 19 inch FC Trinitron CRT monitor.

The breadth of the menu affected the layout of the labelled links for each page. Labelled links were presented in vertical columns in font 15 Comic Sans MS. The approximate distance between each label was three degrees of visual angle. Each column contained approximately eight labelled links, therefore the as the breadth of the menu increased the number of vertical columns also increased. Figure 5.1 provides an example web page from the current experiment.

In the experiment, each trial commenced by the participant first reading the goal statement. When the participant was ready, they selected a search button with the mouse, which then presented the web page on the screen and removed from the screen the goal statement. Participants were instructed to scan through the labelled links on the page and to select the link that they believed to be most relevant to the goal statement as quickly and accurately as possible. In order to impose a meaningful cost structure to the task, participants did not progress to the next trial until they selected the correct link. The location of the target item each page was randomised.

Eye tracking was performed using an ASL Pan/Tilt optics eye tracking system. Eye movement data was sampled at a rate of 50 times per second (once every 20 ms). Eye movement fixations were determined using the Applied Science Laboratories *Eyenal* software package.

5.2.2. Results

Figure 5.1 presents a typical eye movement trace from the current experiment. As can be seen in Figure 5.1, the labelled links were arranged in two vertical columns, each containing 8 items. As before, fixations were mapped to an item in the menu if they landed within the item's respective area of interest. Areas of interest were defined as a standardized rectangular area around each menu item (occurring at the mid-point between vertically contiguous items). Fixations that did not land over a menu item were ignored (accounting for less than 5% of all fixations).

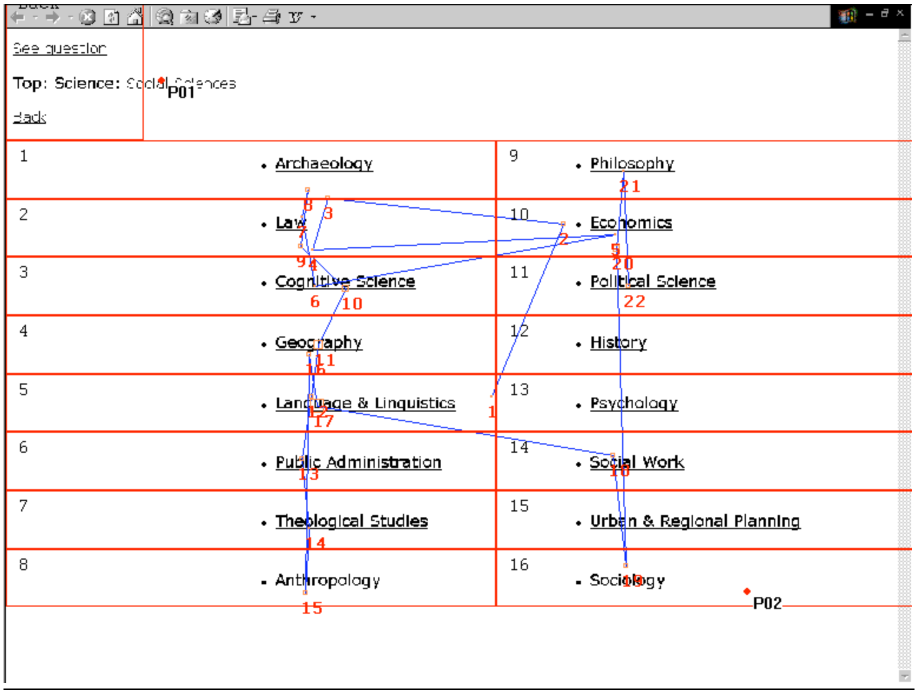


Figure 5.1. A typical eye movement trace where the participant searched a web page containing 16-items and the labels were medium distance from the target. The goal statement was "Find out the capital of South Korea", and the target was the labelled link "Geography". Rectangular boxes around the labelled links define areas of interest.

5.2.2.1. Accuracy

The proportion of trials in which participants selected the correct forward link on each web page was comparably low (M = 51.54%, SD = 25.93%). Participants were less likely to accurately select the correct forward link when the depth of the labels in the web site structure were far from the target (*M* = 37.04%, SD = 20.75%) compared to when they were medium (*M* = 62.04%, *SD* = 25.39%) or close (*M* = 55.56%, *SD* = 25.20%) in distance to the target. A 3 x 3 (menu breadth x distance from the target) repeated-measures ANOVA found a significant main effect of distance from the target on selection accuracy, *F* (2, 22) = 11.037, *p* < .001, *MSE* = .055. Tests of within-subjects contrasts revealed a significant linear trend, *F* (1, 11) = 11.224, *p* = .01, *MSE* = .055, suggesting that participants were less accurate in their selection when the correct forward link was further from the target in the site hierarchy. There was a trend such that selection accuracy tended to decrease as the number of items increased from 8- (*M* = 58.33%, *SD* = 31.24%), 16- (*M* = 47.22%, *SD* = 16.67%), to 32-items (*M* = 49.07%, *SD* = 27.01%). Although the main effect of menu breadth on selection accuracy was not quite statistically significant at the appropriate .05 alpha-level, *F* (2, 22) = 2.866, *p* = .078, *MSE* = .045. The interaction between menu breadth x distance from the target was significant, *F* (4, 44) = 6.230, *p* < .001, *MSE* = .042. Follow-up tests found that selection accuracy decreased as the number of items on the web page increased when the depth of the labels in the web site structure were either medium or far to the target, *F* (2, 22) = 13.556, *p* < .001, *F* (2, 22) = 10.811, *p* = .005, respectively. There was not a significant simple effect of menu breadth on selection accuracy when the labels on the web site were close to the target, *F* (2, 22) = 1.765, *p* = .221.

5.2.2.2. Time to selection

Participants tended to take more time to the selection of an item as the number of items in the menu increased from 8- (*M* = 8.02 s, *SD* = 3.50 s), 16- (*M* = 10.48 s, *SD* = 4.70 s), to 32-items (*M* = 12.80 s, *SD* = 6.09 s). A 3 x 3 (menu breadth x distance from the target) repeated-measures ANOVA found significant main effects on response latency of menu breadth, *F* (2, 22) = 41.669, *p* < .001, *MSE* = 4.983, and the distance from the target, *F* (2, 22) = 18.487, *p* < .001, *MSE* = 7.933. The interaction between menu breadth x distance from the target was also significant, *F* (4, 44) = 5.984, *p* = .005, *MSE* = 10.358. Follow-up tests revealed when the labels in the web site structure were far from the target that there was a significant simple effect of menu breadth on response latency, *F* (2, 22) = 31.677, *p* < .001. There was not a significant simple effect of menu breadth on response latency when the labels in the web site structure were close in distance to the target, *F* (2, 22) = 2.195, *p* = .162, nor medium in distance to the target, *F* (2, 22) = 2.787, *p* = .109.

5.2.2.3. Proportion of items visited/revisited

Analysis of eye movement protocols considered the proportion of the items that were visited (i.e., fixated at least once) prior to selection. A proportional analysis was adopted because the number of labelled links on each web page was manipulated and the target was equally likely to occur in a given quadrant of the page. Figure 5.2 shows the proportion of items that were visited and revisited for each experimental condition.

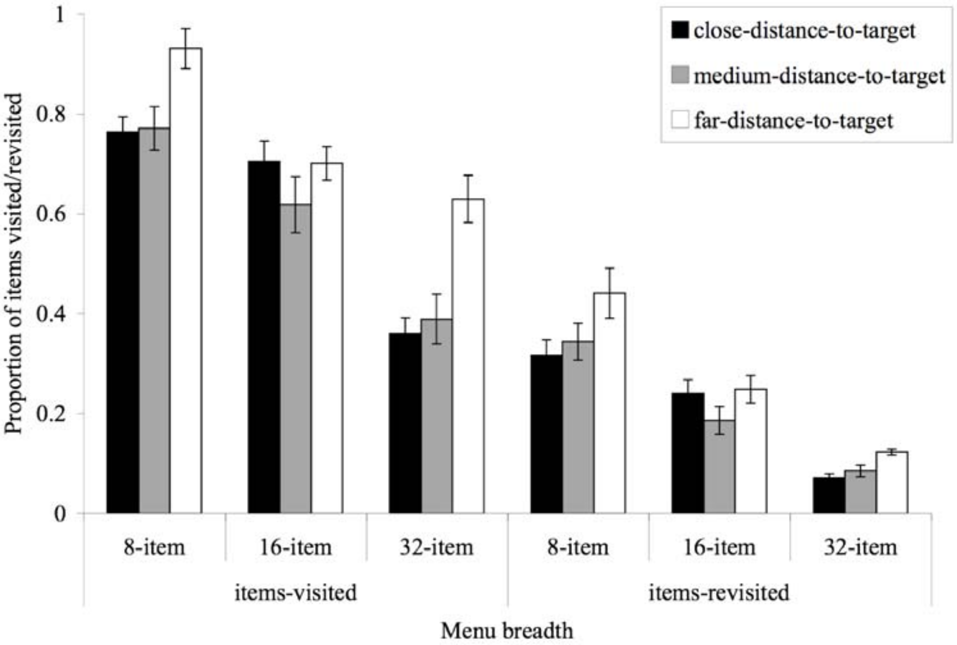


Figure 5.2. Proportion of items visited/revisited across varying menu breadth and distance to the target.

A 3 x 3 (menu breadth x distance from the target) repeated-measures ANOVA found that the proportion of items that were visited at least once was significantly affected by the distance of the labels from the target, *F* (2, 22) = 13.854, *p* < .001, *MSE* = .020, and also the number of labelled links on the web page, *F* (2, 22) = 72.575, *p* < .001, *MSE* = .016. Tests of within-subjects contrasts revealed significant linear trends suggesting that participants visited proportionally fewer items on the web page as the number of items increased, *F* (1, 11) = 169.630, *p* < .001, *MSE* = .014, and also as the distance of the labels from the target increased, *F* (1, 11) = 19.695, *p* = .005, *MSE* = .019. The menu breadth x distance from the target interaction was significant, *F* (2, 44) = 4.116, *p* = .01, *MSE* = .014. Follow-up tests did not alter the main pattern of results, however. There was a significant simple effect of menu breadth on the proportion of items visited at least once, regardless of whether the labels in the web site structure were close, medium, or far from the target, *F* (2, 44) = 139.773, *p* < .001, *F* (2, 44) = 35.942, *p* < .001, *F* (2, 44) = 16.535, *p* = .005, respectively.

Analysis of the proportion of items that were revisited revealed a similar pattern of results as that found for the proportion of items visited at least once. The proportion of items that were revisited was significantly affected by the distance of the labels from the target, *F* (2, 22) = 8.407, *p* = .005, *MSE* = .006, and also the number of labelled links on the web page, *F* (2, 22) = 58.616, *p* < .001, *MSE* = .012. The menu breadth x distance from the target interaction was not significant, *F* (2, 44) = 1.313, *p* = .280, *MSE* = .010. Tests of within-subjects contrasts revealed significant linear trends suggesting that participants revisited proportionally fewer items on the web page as the number of items increased, *F* (1, 11) = 107.392, *p* < .001, *MSE* = .013, and also as the distance of the labels from the target increased, *F* (1, 11) = 8.913, *p* = .05, *MSE* = 1.352.

5.2.2.4. Proportion of first-visit selections

The previous analysis suggests that the number of labelled links on a web page affected the proportion of those items that were visited prior to selection. Indeed, participants were more likely to select an item after visiting for the first time as the number of items increased from 8- (*M* = 21.21%, *SD* = 21.76%), 16- (*M* = 37.37%, *SD* = 24.66%) to 32-items (*M* = 53.54%, *SD* = 29.98%). A 3 x 3 (menu breadth x distance from the target) repeated-measures ANOVA found that the number of labelled links on a web page had a significant effect on the proportion of first-visit-selection searches, *F* (2, 22) = 11.228, *p* < .001, *MSE* = .077. Tests of within-subjects contrasts revealed a significant linear trend, *F* (1, 11) = 13.913, *p* = .005, *MSE* = .124. There was not a significant main effect of the distance of the labels from the target on the proportion of first-visit-selection searches, *F* (2, 22) = 1.579, *p* = .231, *MSE* = .083. The interaction was also non-significant, *F* (2, 44) = 1.367, *p* = .263, *MSE* = .051.

5.2.3. Discussion

The results of Experiment 5 suggest that the number of items on a web page (i.e., menu breadth) and also the depth of the labels within the web site structure (i.e., the distance from the target) had significant consequences for assessment and selection during interactive search. As the number of labelled links on a web page increased participants tended to visit and revisit proportional fewer items prior to selection. This finding was partly because participants were more likely to select an item immediately after visiting it for the first time as the number of options on the page increased. The results of the study are consistent with those from early studies of how people search database menu pages (MacGregor, Lee, & Lam, 1986). Clearly, participants were not visiting all of the items in the available choice set prior to selection and the findings are therefore contrary to assumptions implicit in assess-all accounts of web search (e.g., Pirolli & Card, 1999; Pirolli & Fu, 2003).

A major concern that overshadows the interpretation of the results of Experiment 5 is that label relevance appears to have been confounded within the design. The materials were constructed from labelled links sampled from real web sites and label relevance was not itself manipulated. Selection accuracy tended to decrease as the number of items in the menu increased and also as the depth of the labels within the web site from the target increased. Moreover, the proportion of trials in which participants selected the correct forward link on each web page was comparably much lower (*M* = 51.54%, *SD* = 25.93%) than the results of the experiments in Chapter 3.

Consequently, the results of Experiment 5 cannot directly discriminate between the differing predictions derived from models, such as Miller and Remington (2004) threshold model, Cox and Young's (2004) model, or the ACT-R model described in Chapter 4, of the affect of varying the number of items on a web page for assessment and selection. For instance, the finding that participants were more likely to select an item immediately after visiting it for the first time as the number of options on the page increased is ostensibly consistent with all three accounts, dependent on the assumed distribution of relevance with changing menu breadth. Miller and Remington's threshold model would be supported if the proportion of competing distractor items had also increased with breadth because the probability of encountering an item above threshold, given an assessment of an item, would have increased as the number of items on a page increased. Whereas if the proportion of competing distractors had not increased with increasing breadth, both Cox and Young's model and the ACT-R model would have been supported because in both cases the presence of competing distractors would lead to further assessment when there were fewer items on a page. Moreover, it is not clear whether the reason participants were choosing to select sooner was because the number of items increased or because the number of competing distractor items was affected by increasing menu breadth.

5.3. Experiment 6

Experiment 6 manipulated the number of items in the menu choice set and also the number of competing items in the choice set. This latter manipulation was aimed at addressing some of the concerns over the interpretation of the results of Experiment 5. The primary aim of Experiment 6 was to tease apart model-based predictions of the implications of varying the number of items on a web page for assessment and selection during interactive search. A particular discrepancy described earlier is that Cox and Young's (2004) model predicted that as the number of available items on a given web page increases, selection should be favoured over further assessment when a highly relevant item has been identified. This is because the proportion of information that can be gained through the assessment of a particular item decreases, whereas the cost of an assessment remains the same regardless. In contrast, the ACT-R model, described in Chapter 4, predicted that as the number of available options in the choice set increases selection should be deferred in order for further items to be assessed. This is because as the number of items in the potential choice set increases the attentional focus becomes more distributed amongst those items, therefore, reducing the likelihood that any single item would be initially evaluated as relevant enough to warrant selection. Experiment 6 was designed to discriminate between these predictions by manipulating the number of items in the menu choice set, while the (absolute) position of the target within the list was held more or less constant. In other words, would participants choose to continue assessing items in the menu, after encountering a highly relevant item, simply because more were available to be assessed?

Experiment 6 also manipulated the number of competing items in the menu choice set. A series of predictions can be derived for the potential affect of manipulating the number of items in the menu choice set which compete for selection with the target. As described previously, Miller and Remington's (2004) threshold account assumes that items on a web page are evaluated until one is encountered with an evaluation greater than some arbitrary selection threshold. Miller and Remington's model predicts that as the number of items on a page that are highly relevant to the goal statement increase, the probability of an item exceeding the selection threshold should increase. Participants should therefore be more likely to select an item after visiting it for the first time when there are more competing items. In contrast, the ACT-R model predicts that participants should be less likely to select an item after visiting for the first time when there are more competing items because the probability of any single item exceeding the threshold is decreased. Cox and Young's model makes a similar set of predictions. This is because when there are more items that are relevant to the goal there is greater uncertainty which item should be selected, there further assessment is worthwhile as it reduces this uncertainty.

The above outlines the potential local influence of manipulating the number of competing items in the menu choice set. The results of Experiment 4 suggested that the relevance of labels in the previously experienced (or distant) choice set influences subsequent search behaviour. Recall that models of interactive search can be refined to account for this finding (Cox & Young, 2004; Fu, in press). Theoretically these accounts assume people are sensitive to the experienced distribution of relevance over subsequent searches. An alternative explanation might be that participants in Experiment 4 were strategically learning to favour or disfavour a strategy for immediate selecting attractive items. Such an alternative account is consistent with theoretical explanations of how people choose between competing operators during problem solving (e.g., Lovett, 1998; Lovett & Anderson, 1996). Experiment 6 provided an opportunity to distinguish between these theoretical interpretations. In the experiment, the number of competing items was manipulated and participants either searched menus containing a single target item (i.e., where the competing items were distractors) or many target items (i.e., where competing items were targets). In this way the history of experienced relevance was equivalent between participants, but there was a potential divergence in the history of selection accuracy.

5.3.1. Method

5.3.1.1. Participants

Sixteen Cardiff University undergraduate psychology students participated in return for course-related credit. None of the participants had taken part in any of the other experiments reported in this thesis. All participants were native English speakers and had normal uncorrected vision. All participants were experienced in using a World Wide Web browser, and all had been required to use various computer software packages to produce coursework.

5.3.1.2. Design

The experiment was a 3 x 2 x 2 (number of items x number of competing items x type of competing items) mixed design. During the experiment all participants searched pages that contained 7-, 14-, or 21-items. In addition to the target item there were either two or four additional highly relevant items that served as competing items. The experiment manipulated whether these competing items were either targets (i.e., lead to the goal) or distractors (i.e., did not lead to the goal). The type of competing item was manipulated as a between-subjects variable. As in the previous experiments, eye-tracking data of participants' eye movements up to and including the first selection of an item was of primary interest.

5.3.1.3. Materials and procedure

As in previous experiment, participants completed 28 interactive search tasks (four practices trials followed by four trials for each of the experimental conditions), which required them to search a simplified web page for information relevant to a given goal statement. The goal statements were the same as those in the experiments from Chapter 3 (Experiment 1, 2, & 3). Participants' ratings of sampled web-labels (see Experiment 2 for details) allowed menus containing 7-, 14-, or 21-items to be generated. Each menu contained a single target item that was rated as highly relevant to the goal (i.e., received a median rating of 1 from participants). The position of the target item was held constant across different menu breadths. The target was always located between positions 3 – 6 in the list of labelled links. In addition to the target item there were either two or four additional highly relevant items (i.e., receiving a median rating of 1 or 2 from participants) that served as competing items. The remaining items in the menu were not relevant to the search goal (i.e., they received a median rating of 5 from participants).

In the experiment, the competing items were either target items or distractor items. If a competing item was a distractor then when selected the participant was informed that they had made an incorrect selection and did not progress to the next trial. Whereas, if a competing item was a target then when selected the participant was informed that they had made a correct selection and progressed to the next trial, even if they did not select the single, designated target item.

The experimental materials were controlled by a purpose-built Microsoft Visual Basic program presented on a high contrast 19 inch FC Trinitron CRT monitor. The menu items were presented in a vertical list in font 15 Cosmic Sans MS and the approximate distance between each label was three degrees of visual angle. Each trial commenced by the participant first reading the goal statement. When the participant was ready, they selected a search button with the mouse, which then presented the menu and removed the goal statement from the screen. Participants were instructed to scan through the menu items, commencing their search at the top of the menu, and to select the item that they believed to be most relevant to the goal statement as quickly and accurately as possible. As before, in order to impose a meaningful cost structure to the task, participants did not progress to the next trial until they selected the target item (including additional competing distractor items when these acted as target items). Eye movement data was recording using an ASL Pan/Tilt optics eye tracking system, which was sampled at a rate of 50 times per second. Eye movement fixations were determined using the same procedure outlined in Experiment 5. Experimental trials were blocked by menu breadth, such that participants completed all trials containing a given number of items.

5.3.2. Results

5.3.2.1. Subjective feedback of accuracy

In the current study, a 3 x 2 x 2 (number of items x number of competing items x type of competing items) mixed design ANOVA was employed. The type of competing distractor item (i.e., whether they acted as target or distractor items) was a between-subjects factor. Although selection of the *actual* target item did not differ depending on whether the competing items were targets (*M* = 64.06%, *SD* = 27.24%) or distractors (*M* = 65.62%, *SD* = 21.65%), *F* (1, 14) = .368, *p* = .554, *MSE* = .016, participants received feedback that they almost always made a correct selection when the competing items were targets (*M* = 98.95%, *SD* = 5.05%). Whereas, when the competing items were distractors, participants often selected one of those items instead of the designated target item (*M* = 65.63%, *SD* = 21.65%). This difference in subjective feedback of accuracy was statistically significant, *F* (1, 14) = 238.933, *p* < .001, *MSE* = .011. Therefore while there was a divergence in the history of selection accuracy, the history of experienced relevance was equivalent between the two groups of participants.

Participants that experienced trials where there were competing distractors items were analysed further to explore the effect of varying the number of items in the menu and the number of competing items on selection accuracy. (When the competing items were additional targets selection accuracy was clearly at ceiling.) Participants experienced greater success with selection when there was only two competing distractors (*M* = 69.79%, *SD* = 22.09%) compared to when there was four competing distractors (*M* = 61.46%, *SD* = 20.82%), *F* (1, 14) = 4.667, *p* = .05. Although there was a significant simple effect of menu breadth, *F* (2, 14) = 7.106, *p* = .01, participants selection accuracy was not clearly affected by varying the number of items in the menu from 7- (*M* = 68.75%, *SD* = 25%), 14- (*M* = 56.25%, *SD* = 14.43%) to 21-items (*M* = 71.88%, *SD* = 22.13%).

5.3.2.2. Time to selection of an item

Participants selected an item sooner when there were 7-items (*M* = 4.73 s, *SD* = 1.53 s) in the menu compared to when there were 14-items (*M* = 8.30 s, *SD* = 2.31 s) or 21-items (*M* = 8.74 s, *SD* = 3.88 s) in the menu, *F* (2, 28) = 40.267, *p* < .001, *MSE* = 3.844, and tests of within-subjects contrasts revealed a significant linear trend, *F* (1, 14) = 50.009, *p* < .001, *MSE* = 5.144. Participants selected an item earlier when there were two competing distractor items (*M* = 6.76 s, *SD* = 2.81 s) in the menu compared to four competing items (*M* = 7.76 s, *SD* = 3.63 s), *F* (1, 14) = 10.286, *p* = .01, *MSE* = 2.344, and tests of within-subjects contrasts revealed a significant linear trend, *F* (1, 14) = 10.286, *p* = .01, *MSE* = 2.344. The number of items x number of competing items interaction was significant, *F* (2, 28) = 25.323, *p* < .001, *MSE* = 1.767. Follow-up tests of the interaction did not reveal any significant deviations from the pattern of results found by tests of the main effects: when there were fewer items in the menu participants selected an item sooner, regardless of whether there was two or four competing items in the menu, *F* (2, 28) = 20.927, *p* < .001 and *F* (2, 28) = 47.374, *p* < .001, respectively. Participants also tended to select item earlier when the competing items were additional targets (*M* = 6.00 s, *SD* = 2.70 s) compared to when they were competing distractors (*M* = 8.52 s, *SD* = 3.32 s), *F* (1, 14) = 9.292, *p* = .01, *MSE* = 16.403. There were non-significant second-order interactions between the type of competing item x the number of items, *F* (1, 14) = .256, *p* = .621, *MSE* = 2.344, and the type of competing item x the number of competing items, *F* (2, 28) = 2.024, *p* = .151, *MSE* = 3.844. The third-order interaction was also non-significant, *F* (2, 28) = .715, *p* = .498, *MSE* = 1.767.

5.3.2.3. First-visit-selection

The critical question addressed in the current study was whether participants would choose to continue assessing items in the menu when more were available. Analyses of eye movement protocols therefore only considered the proportion of trials in which participants selected an item after visiting it for the first time (i.e., whether participants made a first-visit-selection). Figure 5.3 shows that the number of items on the page had a significant affect on the proportion of trials in which participants made a first-visit-selection, *F* (2, 28) = 11.203, *p* < .001, *MSE* = .058. Tests of within-subjects contrasts revealed a significant linear trend, *F* (1, 14) = 7.632, *p* = .05, *MSE* = .080, suggesting that participants were more likely to select an item after visiting it for the first time when there were fewer items in the menu. The proportion of first-visit-selection trials was not significantly affected by the number competing items in the menu, *F* (1, 14) = .067, *p* = .800, *MSE* = .027. Although the number of items x number of competing items interaction was significant, *F* (2, 28) = 6.564, *p* = .005, *MSE* = .049, Figure 5.3 shows that this was possibly spurious, owing to the exceptionally low value for a single experimental condition (i.e., two-competing x 14-item).

Figure 5.4 shows that whether participants searched menus that contained a single target item (i.e., competing items were distractors) or many target items (i.e., competing items were targets) had an affect on first-visit-selections, but only when there were more of those competing items. Indeed, when the competing items were distractors participants were less likely to select an item after visiting it for the first time compared to when those competing items were targets, *F* (1, 14) = 5.495, *p* = .05, *MSE* = .134. The type of competing item x number of competing items interaction was significant, *F* (2, 28) = 8.657, *p* = .05, *MSE* = .027, such that when there were four competing items there was a significant effect of whether those items were additional target items or distractors on first-visit-selections, *F* (1, 14) = 9.771, *p* = .01, *MSE* = .031. When there was only two competing items there was not a significant effect of those items on first-visit-selections, *F* (1, 14) = 1.014, *p* = .331, *MSE* = .023. Finally, the second-order interaction between the type of competing item x the number of items was not significant, *F* (2, 28) = .350, *p* = .707, *MSE* = .058, as was the third-order interaction, *F* (2, 28) = 1.051, *p* = .363, *MSE* = .049.

5.3.3. Discussion

The results of Experiment 6 suggest that participants were more likely to select an item immediately after visiting it for the first time when there were fewer, as opposed to more, items in the menu. This finding is contrary to the results of the Experiment 5, and also previous research within the literature (e.g., MacGregor, Lee, & Lam, 1986). In contrast to previous experiments, the current study controlled the semantic relevance of the labelled links in the choice set to the goal statement. Furthermore, while the number of items on a web page was manipulated, the (absolute) position of the target within the list was held constant. The design of the experiments therefore provided a stringent test of whether participants would choose to continue assessing items in the menu, after encountering a highly relevant item, when more were available to be assessed. The findings support the ACT-R model described in Chapter 4, which predicted that as the number of options in the choice set increased, participants would tend to defer selection, in order assess more of the remaining items in the menu. In contrast, the results of the experiment do not support the predictions derived from Cox and Young’s (2004) model, which assumes that the value of assessment is decreased when there are fewer options in the available choice set. The results of the study do not support Miller and Remington’s (2004) threshold model, which assumes that the decision to select an item is not sensitive to the *number* of options available, but the proportion of highly relevant (i.e., above threshold) items in the choice set.

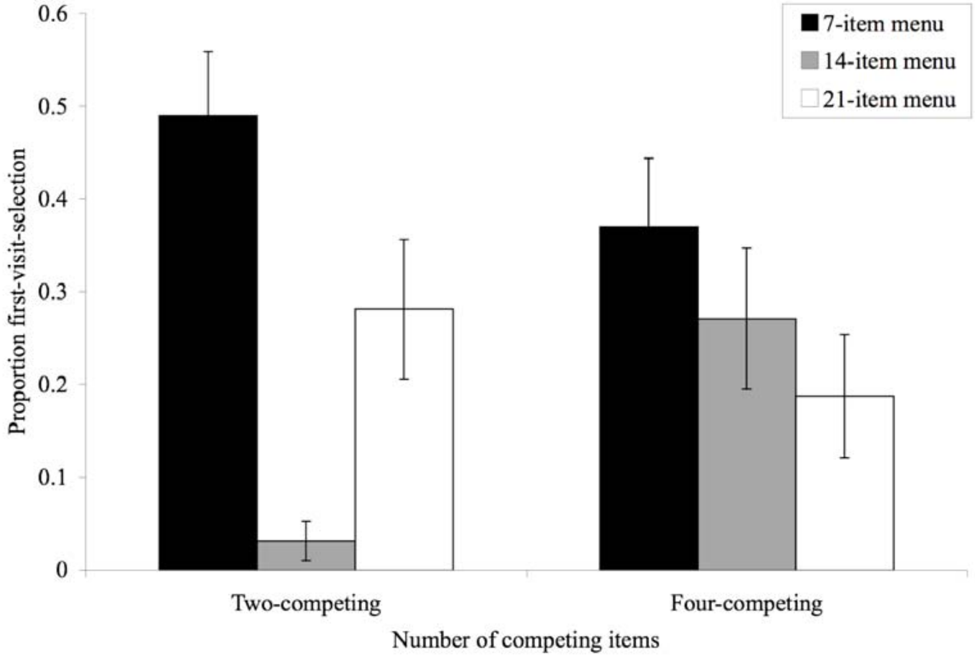


Figure 5.3. **The proportion of trials in which the participants selected an item after visiting it for the first time across varying menu breadth and number of competing items.**

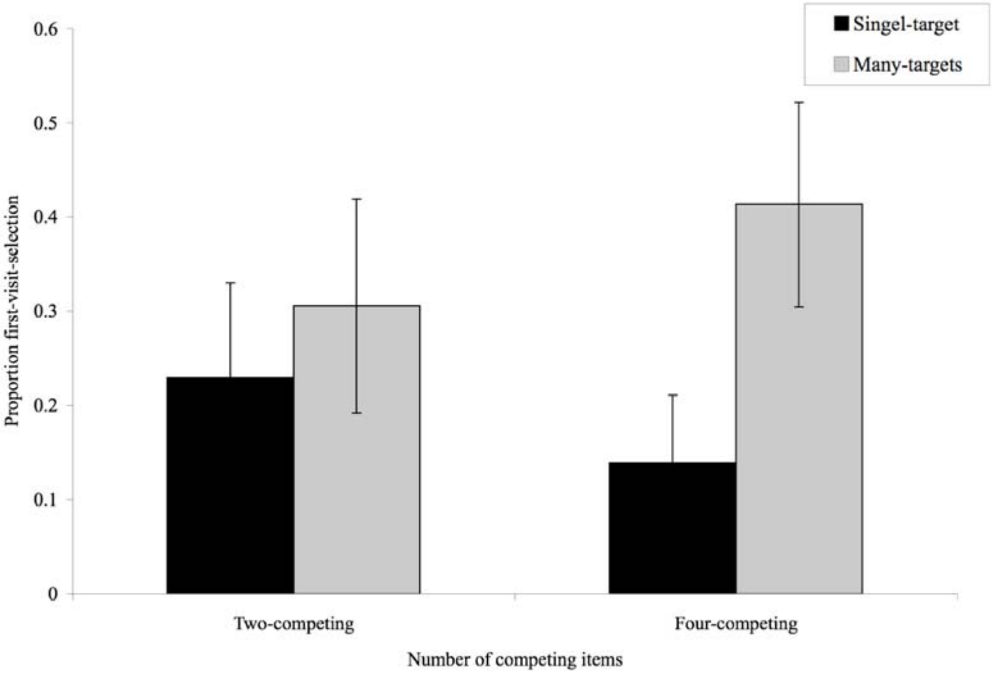


Figure 5.4. **The proportion of trials in which the participants selected an item after visiting it for the first time across varying number of competing distractors and targets.**

Contrary to the predictions of a threshold account (e.g., Miller & Remington, 2004) the results of Experiment 6 found that participants spent more time prior to selection of item when there were more competing (i.e., highly relevant) items in the available choice set. Figure 5.3 suggests that participants were also less inclined select an item after visiting it for the first time when there were more highly relevant items in the choice set (when the exceptionally low value for the two-competing x 14-item experimental condition is disregarded). Moreover, these findings are consistent with those from the earlier experiments in Chapter 3 of this thesis and support context-sensitive accounts (e.g., ACT-R model, Chapter 4; Cox & Young, 2004; Young, 1998) of interactive search which assume that the relative value of assessment increases when there are many items in the available choice set which are relevant to the goal.

It was previously demonstrated in Experiment 4 that the discriminability of the target item affects subsequent search behaviour. Recall that it was unclear in the previous study whether participants were sensitive to the correctness of their selection (i.e., sensitive to the history of successful selection, Lovett & Anderson, 1996) or simply the history of experienced relevance (Fu, in press; Young, personal communication). The current study found that when the selection of a highly relevant item frequently did not lead to the goal, participants did not tend to make an immediate selection on subsequent trials, but opted to further assess items in the menu. Whereas, if the selection of a highly relevant item lead to the goal, then participants tended to favour early selection on subsequent trials. This finding demonstrates that people are at least sensitive to the history of successful selection independently of the history of experienced relevance. It is not clear how current models of interactive search (Cox & Young, 2004; Fu, in press; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Young, 1998) might account for this finding. Although a promising avenue for further work might be within the ACT-R framework (Anderson et al., 2004; Anderson & Lebiere, 1998) as the production rule learning mechanism in the architecture is well suited to capturing the influence of past experience on operator selection over consecutive trials (Lovett & Anderson, 1996).

Finally, a lingering concern in the interpretation of the finding that participants were more likely to make a first-visit-selection when there were fewer items in the available choice set is that as the number of items in the menu increases proportionally fewer of those items would have been assessed when the target was initially encountered. The results of Experiment 3 found that participants were more likely to make a first-visit-selection of an item when more of the items in the choice set had been assessed. Therefore, participants may not have been less likely to select an item after visiting it for the first time just because there were more items in the available choice set, but because there were less unassessed items in the choice set. Furthermore, both of the above interpretations of the empirical data are consistent with predictions derived from the ACT-R model. Further work is required.

5.4. General Discussion

In summary, the experiments in Chapter 5 attempted to tease apart predictions derived from previous models of interactive search (e.g., ACT-R model, Chapter 4; Cox & Young, 2004; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003) of the affect of varying the number of labelled links on a single web page for assessment and selection. Experiment 5 found that as the number of labelled links on a web page increased participants tended to visit and revisit proportional fewer items prior to selection, which was partly because participants were more likely to select an item immediately after visiting it for the first time. A major concern that overshadowed the interpretation of the results of Experiment 5 was that label relevance was not controlled and may have been confounded across varying menu breadths. Experiment 6 dealt with this concern by controlling relevance, while manipulating the number of items in the menu. Furthermore, while the number of items was manipulated, the absolute position of the target within the list was held constant, providing a test of whether participants would choose to continue assessing items because more were available. Consistent with the ACT-R model described in Chapter 4, Experiment 6 found that as the number of available options in the choice set increased participants tended to defer selection in order assess more of the remaining items. This finding was inconsistent with the predictions derived from Cox and Young's (2004) model and also Miller and Remington's (2004) threshold account.

Although the experiments in Chapter 5 were designed to expose the implications of varying the number of labelled links on a single web page for assessment and selection it is not at present clear what the implications are for the generality of the findings are for multiple-page search. In particular, for Experiment 5 the experimental materials sampled labels from web pages at various depths (i.e., the distance from the target) within the structure of the site. It was an open empirical question how the distribution of label relevance was affected by distance from the target. The results of the study found that selection accuracy decreased with distance from the target. This finding implies that the relevance of a label decreases on average with distance from the target node in the web site. Consequently, web sites that opt for a deeper structure are in danger of decreasing target discriminability on top-level pages and therefore impacting the usability of their site. Further empirical investigation is required to accurately characterise the distribution of relevance over an entire web site, as relevance is unlikely to be uniformly distributed.

Finally, the results of Experiment 6 demonstrate that people are also sensitive to the history of successful selection over consecutive searches. When the selection of a highly relevant item frequently did not lead to the goal participants tend to devaluate early selection on subsequent trials in favour of further assessment, whereas, when the selection of a highly relevant item lead to the goal participants tended to favour early selection on subsequent trials. This finding suggests that people do not adopt a fixed and rigid search strategy, but that behaviour is instead highly adaptive to the broader task environment. This is an important consideration for the development of engineering models that aim to provide clear and unambiguous information to support future web design.

CHAPTER 6: SUMMARY AND CONCLUSIONS

6.1. Summary of Results

The goal of the thesis was to investigate strategies for choosing between assessment and selection during interactive search. The major results for each empirical chapter are summarised in turn. Table 6.1 also provides a summary list of the main empirical regularities that were found.

Chapter 3 presented a series of experiments that were designed to discriminate between the three main views of the strategies people adopt during interactive search, namely assess-all, threshold or context-sensitive selection strategy. A series of experiments systematically manipulated the relevance and location of items and measured eye fixations up until selection. These measures were used to calculate the number of visits made to each item, and in turn infer which items were assessed. Experiment 1 was an exploratory study, which attempted to use an automated tool (LSA), to determine label relevance in order to set up experimental conditions. Experiment 1 was considered a failed experiment, in the sense that, for a number of reasons discussed earlier, LSA did not provide a reliably accurate measure of relevance. Consequently, human relevance judgements were used throughout the remainder of the empirical studies. Results from Experiment 2 found that participants visited and revisited fewer items prior to selection when the items in the available choice set were less relevant to the goal. Indeed, Experiment 3 found that participants were more likely to select an item without further visits after more items in the local choice set had already been visited, especially when those items previously visited were less relevant to the goal. These findings support context sensitive accounts of interactive search (Cox & Young, 2004; Young, 1998) but do not support the hypothesis that people persistently assess every item prior to deciding which to select (Pirolli & Card, 1999; Pirolli & Fu, 2003), nor for the idea that people assess items until the most recent exceeds a threshold (MacGregor, Lee & Lam, 1986; Miller & Remington, 2004). It was also found that when a goal-relevant item is located, participants sometimes choose to check the remaining items in the menu, but are more likely to skip some of these items. Experiment 4 found that when the selection of a highly relevant item frequently did not lead to the goal participants learnt to devaluate early selection on subsequent trials in favour of further assessment.

Chapter 4 explored computational models of the choice between assessment and selection during interactive search and whether these models can account for the main experimental results reported in the previous chapter. Context-sensitive models of interactive search rely on a *normalization* assumption (Young, 1998) in order for the subjective value of selecting an item to be sensitive to the context provided by the previously visited item in the choice set. A cognitive model of interactive search was developed that was inspired by Young’s normalization assumption, but which was also sensitive to the psychological constraints encoded in the ACT-R theory of the human cognitive architecture (Anderson et al., 2004). In the model, the evaluation of a label’s relevance was in part dependent on the attentional focus provided by the context of the previously assessed items, therefore, the item that eventually exceeds the selection threshold gained activation through the assessment of these alternative choices. Consistent with the empirical data from Chapter 3, the model rarely visited all of the items in the available choice set prior to the selection of an item and was more likely to select the target item when the surrounding distractor items were less relevant to the goal statement. The model was also more likely to select the target immediately when it is encountered later rather than earlier during the search. The model was inconsistent with the data in at least one respect; it did not provide a good fit to proportion of skipping downward gaze transitions that occurred between spatially non-contiguous items. The model would skip spatially contiguous items when it chooses to reassess items, however, the data from Experiment 3 found that participants were more likely skip items when the target was located towards the top of the menu, compared to when the target was located at the bottom of the menu. A novel exploration of the space of the behaviour of Cox and Young's (2004) model revealed that this shift in how people choose to assess items in the menu might be a rational adaptation to the task environment in which people back-off to a lower cost, lower quality assessment procedure following the assessment of a highly relevant item.

The empirical findings and the model developed in the previous chapters suggest that varying the number of labelled links on a page might have consequences for the choice between assessment and selection. This was a particular pertinent question because one area in which models of interactive search have been applied is to predict the optimal structure of a web site (i.e., whether to prefer broad and shallow or deep and narrow choice sets in a web site design). An engineering model was developed by Miller and Remington (2004), which given a description of the labels and their connectivity, predicted the average length of time for a person to find the information that they require. The model assumed that search of each page was based on a simple threshold strategy, which has been shown to fail to capture performance on single page search tasks. The experiments in Chapter 5 were designed to discriminate between different models-based predictions of the affect of varying the number of labelled links on a page for assessment and selection. Consistent with the ACT-R model described in Chapter 4, Experiment 6 found that as the number of available options in the choice set increased participants tended to defer selection in order assess more of the remaining items. This finding was inconsistent with the predictions derived from Cox and Young's (2004) model and also Miller and Remington's (2004) threshold account. Moreover, the results of the experiments suggest that engineering models, such as that proposed by Miller and Remington (2004), need to be updated so that they are sensitive to the features of the context, such as the distractor semantics and number of distractors.

Table 6.1. Summary of main empirical regularities

Study	Manipulation	Main Findings
Experiment 1	Distractor relevance[1]	Failed experiment
Experiment 2	Target and distractor relevance	1. Fewer items were assessed prior to selection when the target was highly relevant to the goal statement, and also when the distractors were less relevant to the goal statement.
Experiment 3	Target position and distractor relevance	2. More likely to skip over neighbouring items on immediately successive fixations, when the target was located towards the top of the menu. 3. More likely to select the target item immediately after visiting for the first time, when the target was located towards the bottom of the menu and the distractors were less relevant to the goal statement.
Experiment 4	Distractor relevance and difficulty of previous trials	4. More likely to select an item immediately after visiting it for the first time, when previous experience was that selection was more likely to lead to success (i.e., because menu choice sets did not contain competing distractors).
Experiment 5	Number of options and distance from target	5. Proportionally fewer items from the choice set assessed prior to selection as the number of options increases. 6. Proportionally fewer items from the choice set assessed when the sampled labels were closure to the target node.
Experiment 6	Number of options, distractor relevance, and difficulty of previous trials	7. Immediate selection of the target item preferred over further assessment when there were fewer items in the available choice set.

Note 1. For Experiment 1 the label relevance for experimental materials was defined using an automated tool, LSA. For all other studies, however, human relevance judgements were collected in order to define label relevance.

The remainder of Chapter 6 provides a general discussion of the assumptions underlying the use of eye-tracking methodology to support the main conclusions of this thesis, the implications of the main findings for design, and possible directions for future research.

6.2. Eye-tracking Methodology

Analysis of eye movement protocols served as a primary dependent variable in all the experiments presented in this thesis. These measures were used to calculate the number of visits made to each item in the menu, and in turn infer which items were assessed. A central underlying assumption was that gaze shifts were tightly coupled with the allocation of visual perception and cognition, in this case representing the assessment of a labelled link. It is important to note, however, that the main conclusions of this thesis rest only on a liberal interpretation of this assumption. For instance, a central claim made in this thesis is that people made fewer assessments of items when the items in the available choice set were less relevant to the goal. This claim was supported by the fact that participants in the experiments were found to be more likely to select an item after visiting it for the first time when the distractors were of lower quality, therefore implying that fewer links were assessed.

6.3. Implications for Design

The findings are relevant to a range of engineering techniques for predicting the usability of web sites (Blackmon et al., 2002, 2003; Chi et al., 2003; Church & Keane, 2004; Kahr & Hornof, 2005; Pirolli & Fu, 2003). A common feature of these techniques is that they assume that when navigating the web people will select the link on a page with the highest relevance. The underlying analysis assumes that people consider all of the options available on the page. For instance, Chi et al.'s (2003) Bloodhound project provides a usability tool for web designs. The tool takes as an input a particular set of goal statements and then crawls a web site to predict, amongst other usability metrics, the percentage of users that are likely to select each labelled link on the site for a given goal. The findings of the experiments in Chapter 3 and Chapter 5 demonstrate that people may not always visit every item on a web page and that they sometimes select an item prior to having assessed all of the available items. Consequently, techniques which assume that when navigating the web people select the link on a page with the highest relevance may lead to spurious predictions because it is possible for people to miss the item with the highest relevance because they have already opted to select an item that appeared to be good enough to warrant selection.

6.4. Future Directions

The empirical studies presented in this thesis were limited, in some respect, to dealing with cases of search on a single web page (or menu). One potential future direction to extend this work could be to examine eye movement protocols during search through multi-page sites. There are technical difficulties, however, that would need to be overcome in order to hook eye movement protocols to a dynamic display. If this technical difficulty with the eye tracking methodology can be overcome, then the theoretical benefits would be considerable, not least because they would yield further potential developments in relation to the modelling of interactive search. For example, what are the implications for the relative cost of assessment and selection embedded within navigation choices (i.e., between selecting a forward link or choosing to backup to a previously visited page)?

There is at least some evidence to suggest that the decision to leave a page (i.e., to select a backup button) is influenced by memory for the quality of the unselected item on a previously explored page (Howes, Payne, & Richardson, 2002). Such findings would at present be difficult to integrate with context-sensitive models of interactive search particular as extending a rational analysis of interactive search to multiple-page search has proven difficult (Young, personal communication). Extending the ACT-R model described in Chapter 4 to multi-page search would be an interesting avenue to explore. One idea might be to further exploit the ACT-R theory of declarative memory (Anderson & Lebiere, 1998) to incorporate Howes, Payne, and Richardson's (2002) model of using an episodic memory of previous assessments to guide navigational choices. Such a revision to the model could potentially have consequences for how candidate items are explicitly represented in declarative memory.

A further area that is ripe for empirical investigation is the generality of the findings for interactive search that takes place on small screen devices, such as cellular phones and PDA's. For instance, cellular phones offer a range of functionality other than simply making calls (including tools for managing contact information, voice mail, hardware settings, and often software for playing games and browsing the Web) that is often accessed through a menu structure. In terms of the classic depth vs. breadth trade-off, there is at least some empirical evidence to suggest that a broad navigation structure, which as discussed previously was found to be superior during search on personal computers, also has an advantage for a small-screen device such as the cellular phone (Parush & Yuviler-Gavish, 2003). This finding is interesting given that an implication for cellular phones, compared to traditional personal computers, is that interactions techniques are generally more costly, requiring discrete selections actions in the form of button presses. It is known that information acquisition behaviour is influenced by the cost of accessing information from the environment (Lohse & Johnson, 1996). Further empirical work is required to determine whether strategies for exploring the menu structure on a cellular phone differ from those on a personal computer that have lower interaction costs.

In terms of modelling, there has been some progress in developing models to support menu design on cellular phones (St. Amant, Horton, & Ritter, 2004). These might benefit from the developments in modelling the processes underlying single-page explored in this thesis. For instance, the attentional focusing mechanism employed in the ACT-R model described in Chapter 4 might be pertinent in this context.

Finally, it is also worth commenting on the relationship of the work presented to more general web-based activities, such as searching the results list of a search engine (e.g., Google, MSN Search, & Yahoo). Search engines are of particular interest at the moment because they provide a powerful tool to support users' goal-directed search of the web. Clearly, there are important differences between the results page of a search engine and the type of interactive search task described here. For instance, search engines are designed to return a fairly homogenous set of results clustering around the query terms, often with content rich abstracts, whereas labels on a web page aim to provide discriminatory navigation cues with minimal content information for users with often different goals. Nonetheless, the results of a series of recent eye-tracking studies (Granka, Joachims, & Gay, 2004), which examined how users interact with the results page of a search engine, bear some similarities to the findings of the experiments in this thesis. In particular, Granka, Joachims, and Gay (2004) found that people rarely assessed all of the results prior to selection. It was also clear that the top few items receive most attention, with the average fixation time spent on an abstract dropping of sharply after only the second link.

6.5. Conclusion

In summary, the work presented in this thesis has demonstrated that when people search a novel web page for information that is relevant to the achievement of their search goal, they do not simply select the link on a page with the highest scent, nor do they simply assess items until the most recent exceeds a threshold. The empirical studies presented in this thesis demonstrated, that people are in fact more strategic and sensitive to context than previous models suggest. It was found that people sometimes choose an item that appears good enough, but sometimes choose to continue checking. The decision to select an item was found to be sensitive to the relevance of, and also the number of, labelled links available in the immediate choice set. The relevance of the labels in the distant (or previously experienced) choice sets also influenced participants subsequent search behaviour. These findings imply that during interactive search, decisions are continually made about whether to select one of the assessed items immediately or to make further assessments. Each decision is sensitive to the estimated relevance of all of the items so far assessed, and not just to the most recent item.

The theoretical work of modelling these key empirical findings, allows for explicit theories of interactive search (e.g., Cox & Young, 2004; Howes, Payne, & Richardson, 2002; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli & Fu, 2003; Rieman, Young, & Howes, 1996; Young, 1998) to be formulated and evaluated. A model of interactive search was proposed, in which the decision to select an item was governed by a dynamic threshold. The model selected items that exceeded a threshold, but activation was gained through the assessment of alternative choices in the menu. This behaviour is in part, a consequence of a general theory of human declarative memory (Anderson et al., 2004).

The work presented in this thesis is relevant to a range of engineering techniques for predicting the usability of web sites. The development of explicit, formal cognitive models of how people search web sites holds the potential to provide clear and unambiguous information to support future web design and improve usability. In particular, previous models have been developed to predict the time required by a typical user to search a web page structure, or which labelled link a user is likely to select for a given information goal. Towards this end, there has been significant progress within the field (Blackmon et al., 2002, 2003; Chi et al., 2003; Church & Keane, 2004; Kahr & Hornof, 2005; Pirolli & Fu, 2003). The findings presented in this thesis, point to future developments, in which current models might be updated so that they are sensitive to the extent to which people adapt strategy to the features of the context, such as the distractor semantics and number of distractors.

7. REFERENCES

Aaltonen, A., Hyrskykari, A., & Räihä, K.-J. (1998). 101 spots, or how do users read menus? In C.-M. Karat, A. Lund, J. Coutaz, & J. Karat (Eds.), *Proceedings of the ACM CHI 1998 Human Factors in Computing Systems Conference* (pp. 132–139). New York, NY: ACM Press.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillside, NJ: Lawrence Erlbaum.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval. *Psychological Science, 15*, 225–231.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*, 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum.

Budiu, R., & Anderson, J. R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science, 28*, 1–44.

Blackmon, M. H., Kitajima, M., & Polson, P. G. (2005). Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. In G. Veer, & C. Gale (Eds.), *Proceedings of the ACM CHI 2005 Human Factors in Computing Systems Conference* (pp. 31–40). New York, NY: ACM Press.

Blackmon, M. H., Kitajima, M., & Polson, P. G. (2003). Repairing usability problems identified by the cognitive walkthrough for the web. In G. Cockton, & P. Korhonen (Eds.), *Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference* (pp. 497–504). New York, NY: ACM Press.

Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the web. In L. Terveen (Ed.), *Proceedings of the ACM CHI 2002 Human Factors in Computing Systems Conference* (pp. 463–470). New York, NY: ACM Press.

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies, 55,* 41–84.

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. In M. W. Altom, & M. G. Williams (Eds.), *Proceedings of the ACM CHI 1999 Human Factors in Computing Systems Conference* (pp. 402–409). New York, NY: ACM Press.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled web we wove: A taskonomy of www use. In M. W. Altom, & M. G. Williams (Eds.), *Proceedings of the ACM CHI 1999 Human Factors in Computing Systems Conference* (pp. 544–551). New York, NY: ACM Press.

Card, S. K., Pirolli, P., Van Der Wege, M., Morrision, J. B., Reeder, R. W., Schraedley, P. K., & Boshart, J. (2001). Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability. In M. Beaudouin-Lafon, & R. J. K. Jacob (Eds.), *Proceedings of the ACM CHI 2001 Human Factors in Computing Systems Conference* (pp. 498–505). New York, NY: ACM Press.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the world wide web. In P. H. Enslow, & D. Kroemker (Eds.), *Proceedings of the Third International World-Wide Web Conference on Technology, Tools and Applications (*pp. 1065–1073). New York, NK: Elsevier.

Chi, E. H., Rosien, A., Suppattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J., & Cousins, S. (2003). The Bloodhound project: Automating discovery of web usability issues using the InfoScent simulator. In G. Cockton, & P. Korhonen (Eds.), *Proceedings of the ACM CHI 2003 Human Factors in Computing Systems Conference* (pp. 505–512). New York, NY: ACM Press.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions on the web. In M. Beaudouin-Lafon, & R. J. K. Jacob (Eds.), *Proceedings of the ACM CHI 2001 Human Factors in Computing Systems Conference* (pp. 490–497). New York, NY: ACM Press.

Chi, E. H., Pirolli, P., & Pitkow, J. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. In T. Turner, G. Szwillus, M. Czerwinski, F. Peterno, & S. Pemberton (Eds.), *Proceedings of the ACM CHI 2001 Human Factors in Computing Systems Conference* (pp. 161–176). New York, NY: ACM Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Cockburn, A., & McKenzie, B. (2001). What do web users do? An empirical analysis of web use. *International Journal of Human-Computer Studies, 54*, 903–922.

Cox, A.L., & Young, R.M. (2004). A rational model of the effect of information scent on the exploration of menus. Poster session at the meeting of the *6th Internal Conference on Cognitive Modelling*, Pittsburgh, PA.

Farahat, A., Pirolli, P., & Markova, P. (2004). Incremental methods for computing word pair similarity (TR-04-6-2004). Palo Alto, CA: Palo Alto Research Center Incorporated.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford, UK: Oxford University Press.

Franzke, M. (1995). Turning research into practice: Characteristics of display-based interaction. In I. R. Katz, R. L. Mack, L. Marks, M. B. Rosson, & J. Nielson (Eds.), *Proceedings of the ACM CHI 1995 Human Factors in Computing Systems Conference* (pp. 421–428). New York, NY: ACM Press.

Franzke, M. (1994). Exploration and experienced performance with display-based systems *(*Doctoral Dissertation, University of Colorado, 1994). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 56*(2-B), 1134.

Fu, W.-T. (in press). Adaptive Tradeoffs between Exploration and Exploitation: A Rational-Ecological Approach. In W. D. Gray (Ed), *Integrated Models of Cognitive Systems*, Oxford, UK: Oxford University Press.

Fu, W.-T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science, 28,* 901–935.

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. Poster session presented at the *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Sheffield, UK.

Halverson, T., & Hornof, A. J. (2004). Link colors guide a search. Paper presented at *CHI 2004 extended abstracts on Human Factors in Computing Systems.* Vienna, Austria.

Hornof, A.J. (2004). Cognitive strategies for the visual search of hierarchical computer displays. *Human-Computer Interaction, 19*, 183–223.

Howes, A. (1994). A model of the acquisition of menu knowledge by exploration. In B. Adelson, S. Dumais, J. S. Olson (Eds.), *Proceedings of the ACM CHI 1994 Human Factors in Computing Systems Conference* (pp. 445–451). New York, NY: ACM Press.

Howes, A., Payne, S. J., & Richardson, J. (2002). An instance-based model of the effect of previous choices on the control of interactive search. In W. D. Gray, & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 476–481)*,* Mahwah, NJ: Lawrence Erlbaum.

Jansen, B. J., & Pooch, U. (2000). A Review of Web Searching Studies and a Framework for Future Research. *Journal of the American Society of Information Science and Technology, 52*, 235–246.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras, & M. A. Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 151–182). Hillsdale, NJ: Lawrence Erlbaum.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

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

Kieger, J. L. (1984). The depth/breadth trade-off in the design of menu-driven user interfaces. *International Journal of Man-Machine Studies, 20,* 365–369.

Kieras, D.E., & Meyer, D.E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12,* 391–438.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(*2)*, 211–240.

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science, 7,* 161–165.

Lau, T., & Horvitz, E. (1999). Patterns of search: analyzing and modeling Web query refinement. In J. Kay (Ed.), *Proceedings of the Seventh international Conference on User Modeling* (pp. 119–128). Secaucus, NJ: Springer-Verlag.

Larson, K., & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent for information retrieval. In C.-M. Karat, A. Lund, J. Coutaz, & J. Karat (Eds.), *Proceedings of the ACM CHI 1998 Human Factors in Computing Systems Conference* (pp. 25–32). New York, NY: ACM Press.

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

Lesk, M. E., & Salton, G. (1968). Relevance assessments and retrieval system evaluation. *Information Storage Retrieval, 4*, 434–359.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science, 4*, 6–14.

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes, 68*, 28–43.

Lovett, M. C. (1998). Choice. In J. R. Anderson, & C. Lebiere (Eds.), *The atomic components of thought (*pp. 255–296). Mahwah, NJ: Lawrence Erlbaum.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences of operator selection. *Cognitive Psychology, 31,* 168–217.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54 (Whole No. 248).*

Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of Behavior*. Eugene, Oregon: University of Oregon Books.

MacGregor, J., Lee, E., & Lam, N. (1986). Optimizing the structure of database menu indexes: A decision model of menu search. *Human Factors, 28,* 387–399.

McCarthy, J. D., Sasse, M. A., & Riegelsberger, J. (2003). Could I have the menu please? An eye tracking study of design conventions. In E. O'Neill, P. Palanque, & P. Johnston (Eds.), *People and Computers XVII – Designing for Society* (pp. 401–414). London, UK: Springer-Verlag.

Mat-Hassan, M., & Levene, M. (2005). Associating search and navigation behavior through log analysis. *Journal of the American Society for Information Science and Technology, 56*, 913–934.

Miller, C. S., & Remington, R. W. (2004). Modelling information navigation: Implications for information architecture. *Human-Computer Interaction, 19*, 225–271.

Miller, D. P. (1981). The depth/breadth tradeoff in hierarchical computer menus. *Proceedings of the Human Factors Society 25th Annual Meeting (*pp. 296–300). Santa Monica, CA: HFES.

Morrison, J. B., Pirolli, P., & Card, S. K. (2001). A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. Poster session presented at *CHI 2001 Extended Abstracts on Human Factors in Computing Systems Conference*, Seattle, WA.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nielson, J. (1999). *Designing web usability: The practice of simplicity*. Thousand Oaks, CA: New Riders Publishing.

Nilsen, E. L. (1991). Perceptual-motor control in human-computer interaction. Unpublished technical report, University of Michigan, Ann Arbor, MI.

Norman, K. L. (1991). *The psychology of menu selection: Designing cognitive control of the human/computer interface*. Norwood, HJ: Ablex.

Parush, A., & Yuviler-Gavish, N. (2004). Web navigation structures in cellular phones: The depth/breadth trade-off issue. *International Journal of Human Computer Studies, 60 (5-6)*, 753–770.

Payne, S. J., Howes, A., & Reader, W. R. (2001). Adaptively distributing cognition: A decision-making perspective on human-computer interaction. *Behaviour and Information Technology, 20,* 339–346.

Payne, S. J., Richardson, J., & Howes, A. (2000). Strategic use of familiarity in display-based problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1685–1701.

Pearson, R., & van Schaik, P. (2003). The effect of spatial layout of and link colour in web pages on performance in a visual search task and interactive search task. *International Journal of Human-Computer Interaction, 59*, 327–353.

Pierce, B. J., Parkinson, S. R., & Sisson, N. (1992). Effects of semantic similarity, omission probability and number of alternatives in computer menu search. *International Journal of Man-Machine Studies, 37*, 653–677.

Pirolli, P. (in press). The use of proximal information scent to forage for distal content on the world wide web. In A. Kirlik (Ed.) *Working with Technology in Mind: Brunswikian Resources for Cognitive Science and Engineering*. Oxford, UK: Oxford University Press.

Pirolli, P. (2005). Rational of information foraging on the Web. *Cognitive Science, 29*, 343–373.

Pirolli, P., & Fu, W-T.F. (2003). SNIF-ACT: A model of information foraging on the world wide web. In P. Brusilovsky, A. Corbett, & F. de Rosis (Eds.), *User Modeling 2003, 9th International Conference on User Modelling (*Vol. 2702, pp. 45-54)*,* London, UK: Springer-Verlag

Pirolli, P., Fu, W.-T., Reeder, R., Card, S. K. (2002). A user-tracing architecture for modelling interaction with the world wide web. In the proceedings of *Advanced Visual Interfaces,* Trento, Italy.

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

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hill.

Reder, L. M. & Schunn, C. D. (1999). Bringing togeather the psychometric and strategy worlds: Predicting adaptivity in a dynamic task. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction theory and application. Attention and Performance XVII*. Cambridge, MA: MIT press.

Rieman, J. (1994). Learning strategies and exploratory behavior of interactive computer users (Doctoral dissertation, University of Colorado). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 55*(10-B), 4470.

Rieman, J., Young, R. M., & Howes, A. (1996). A dual-space model of iteratively deepening exploratory learning. *International Journal of Human-Computer Studies, 44*, 743–775.

Rosenbloom, P. S., Laird, J., & Newell, A. (1993). *The Soar papers: Research on integrated intelligence*. Cambridge, MA: MIT Press.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1,* 201–220.

Salvucci, D. D., & Anderson, J. R. (2001). Automated eye-movement protocol analysis. *Human-Computer Interaction, 16*, 39–86.

Schunn, C.D. & Reder, L.M. (2000). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130*, 59–76.

Sellen, A. J., Murphy, R., & Shaw, K. L. (2002). How knowledge workers use the web. In L. Terveen (Ed.), *Proceedings of the ACM CHI 2002 Human Factors in Computing Systems Conference* (pp. 227–234). New York, NY: ACM Press.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction, third edition*. Reading, MA: Addision-Weasley.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a Very Large Web Search Engine Query Log. *ACM SIGIR Forum, 33*, 6–12.

Snowberry, K., Parkinson, S. R., & Sisson, N. (1983). Computer display menus. *Ergonomics, 26*, 699–712.

Soto, R. (1999). Learning and performance by exploration: Label quality measured by latent semantic analysis. In M. W. Altom, & M. G. Williams (Eds.), *Proceedings of the ACM CHI 1999 Human Factors in Computing Systems Conference* (pp. 418–425). New York, NY: ACM Press.

Spink, A., Bateman, J., & Jansen, B. J. (1999). Searching the web: A survey of EXCITE users. *Internet Research: Electronic Networking Applications and Policy, 9*, 117–128.

St. Amant, R. Horton, T. E., & Ritter, F. E. (2004). Model-based evaluation of cell phone menu interaction. In E. Dykstra, & M. Tscheligi (Eds.), *Proceedings of the ACM CHI 2004 Human Factors in Computing Systems Conference* (pp. 343–350). New York, NY: ACM Press.

Tauscher, L., & Greenberg, S. (1997). How people revisit web pages: Empirical findings and implications for the design of history systems. *International Journal of Human Computer Studies, 47,* 97–138.

Turney, P. D. (2001). Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In L. De Raedt, & P. A. Flach (Eds.) *Proceedings of the 12th European Conference on Machine Learning* (Vol. 2167, pp. 491–502), London, UK: Springer-Verlag.

Young, R.M. (1998). Rational analysis of exploratory choice. In M. Oaksford, & N. Chater (Eds.). *Rational Models of Cognition*. Oxford, UK: Oxford University Press.

Young, R.M., & Cox, A.L. (2000). A new rational framework for modelling exploratory device learning … but does it fit with ACT-R? In *Proceedings of the Seventh Annual ACT-R Workshop and Summer School,* Pittsburgh, PA, 2000.

8. APPENDIX: EXPERIMENTAL MATERIALS FOR EXPERIMENT 2.

The experimental materials used in this thesis formed part of a database of goal statements and labelled links. In the experiments, participants completed ecologically valid interactive search tasks which required them to search a simplified web page (or menu) for information relevant to a given goal statement. The appendix contains, as an example, the experimental materials for Experiment 2.

In order to derive ecologically determined goal statements, a web usage survey was posted to under-graduate students in the School of Psychology at Cardiff University. The web usage survey aimed to identify search queries, which participants in the experiment would typical have used the web for. From the 25 responses to the survey it was possible to determine 45 unique search goals.

The web usage survey also aimed to find out which web sites the respondents had visited whilst searching the web for each goal. The labels from these suggested web sites were then sampled. For example, for the search goal "*check your bank balance*" labels were sampled from various online banking web sites (e.g., http://www.hsbc.co.uk & http://www.natwest.co.uk). In total, some 2,000 individual labels were sample from various web sites for all category types.

In order to put together menu choice sets for the experiments, the relevance of each of the sampled web-page labels in relation to a particular goal statement had to be first determined. Participants completed a ratings questionnaire in return for course-related credit. (None of the participants that took part in the rating study took part in any of the experiments reported in this thesis.) In the ratings study, participants were instructed to estimate the likelihood that selecting a label would lead to the achievement of the goal. Participants made relevance estimates on a 5-point scale, where 1 represented a label that was very relevant to the goal description and 5 represented a label that was not at all relevant to the goal description. To gather the ratings all of the sampled web-labels for a particular goal description were made available at once and participants asked to rate them one-by-one.

The appendix contains, as an example, the experimental materials for Experiment 2. The materials from the remaining experiments reported in this thesis were similar to those reported and are not presented. In Experiment 2, participants searched menus that differed in terms of the semantic relevance of the target item (highly relevant or moderately relevant) and the semantic relevance of the distractor items (moderately relevant, not relevant, or not at all relevant). Presented below are the actual materials searched by participants in the experiment, grouped by experimental condition. For each trial there was a goal statement and 16 labelled links. The labelled links are presented as they were in the experiment, with the exception that the median participant rating has been placed in parenthesis. Participants made relevance estimates on a 5-point scale, where 1 represented a label that was very relevant to the goal description and 5 represented a label that was not at all relevant to the goal description.

8.1. Highly relevant target item x moderately relevant distractor items

A friend is planning to come and visit you in Cardiff. Find a road map of the city, with directions to your house.

Calendar of Events (4)
Community Events (3)
Towns and Villages (1)
Recreation Services (3)
Housing Tenants Survey (4)
City Development (3)
Cardiff Community Strategy (3)
Public Interest (3)
Cardiff Community Strategy (4)
Leisure Centres (3)
Highways and Parks (3)
Heritages Sites (4)
Libraries (3)
Urban Initiatives (3)
Castles (3)
Environment Groups (4)

You are planning a Spring break. Look into booking a cheap flight to New York.

Tourist Information Centres (2)
Lodging (3)
Mountain Biking (5)
Confirm Flight (4)
Baggage (4)
Special Offers (1)
Road Casualties (5)
Penalty Fares (4)
Hire a Car (4)
Check-in information (3)
Car Rental (4)
Boat Trips (4)
European Advisor (4)
Motoring News (4)
Route Planner (5)
Surfing (5)

Find out about the latest interior design styles for decorating your apartment.

Chests (2)
Lawn and Garden (4)
Towels and Robes (3)
Life Style (1)
Christmas Decorations (4)
Spa Baths (3)
Kitchen Accessories (2)
Housekeeping (3)
Spirits (4)
Rugs (2)
Herbal Remedies (4)
Facials (4)
Nutritional Supplements (4)
Vegetarian (4)
Diet (4)
Beauty (4)

Buy some moisturizers and skin care products.

Style and Culture (2)
Prevention and Treatment (4)
Spa Session Prices (3)
Hair Accessories (3)
Body Care Treatments (1)
Ask the Doctor (3)
Thermal Treatment (3)
Bath Linen (4)
Bath (2)
Nutrition (4)
Yoga (4)
Weight Loss (4)
Exercise (4)
Parenting (4)
Fine Wine Centre (5)
Pottery and China (5)

Buy tickets for an upcoming music concert at the Student Union.

This Month (2)
Gossip (4)
Clubbing Guides (3)
Media Centre (3)
Record Labels (4)
Tourism (3)
Leisure (3)
Event Tickets (1)
Reviews (2)
Night Life (3)
Clubs (3)
Records (3)
Galleries (4)
Trailers (4)
Theatres (4)
Advertising (4)

You are approaching the end of your degree course and need to find out the date of the graduation ceremony.

Postgraduate Enrolment (2)
University News (2)
Semester Dates (2)
Clearing (4)
Seminars (4)
Graduation Dates (1)
Admissions (4)
Term Dates (2)
Academic Programs (3)
Student Advice Centre (3)
Continuing Education (3)
Examinations (3)
Departments (4)
Resource Finder (4)
Teaching Strategies (4)
Monitors (4)

8.2. Highly relevant target item x not relevant distractor items

You are planning to go out to the cinema this weekend, but are not sure what films are currently showing. Read reviews of the recent movie releases.

Video Games (4)
Galleries (4)
New Movies (1)
Crafts (4)
Museums (4)
Gardening (4)
Box sets (4)
Song Lyrics (4)
Theatre and Dance (4)
Music Awards (4)
Quizzes (4)
Radio Programs (4)
Prizes (4)
Restaurants (4)
Collective Arts (4)
Clubbing Guides (4)

Look for a pair of suede boots.

Boys Jackets (5)
Watches (4)
Robes (4)
Belts (5)
Sweaters (4)
Shoes (1)
Custom Jewellery (4)
Scarves (4)
Lingerie (4)
Purses (4)
Bracelets (5)
Pyjamas (4)
Sports Coats (4)
Wallets (4)
Suits (5)
Maternity Clothing (5)

Get the latest news headlines on the Iraq crisis.

Films (4)
Weather (5)
Education (4)
News Stories (1)
Jobs (4)
Home and Garden (5)
Commentary (4)
Business (5)
Sports Talk (4)
Minutes of the Cabinet (4)
Asylum (4)
Arts Reviews (4)
Public Statements (4)
International Resource Centre (4)
The Economy (4)
Football (5)

It is your Mother's birthday. Order a large box of chocolates for the occasion.

Christmas Preparation (4)
Sympathy (4)
Plaques (4)
Bouquets (4)
Birthday Collection (1)
Gifts for Baby (4)
Zodiac Bouquets (4)
Perfume (4)
Perfume (4)
Sapphire (4)
Greeting Cards (4)
Baby Gifts (4)
Party Shop (4)
Special Deliveries (4)
Pens (4)
Autumn Flowers (4)

Read all the latest reports from football matches over the last few days.

- Badminton (4)
- Wind Surfing (4)
- Cricket (4)
- Mobile Text Alerts (4)
- Snow Boarding (4)
- Swimming (4)
- Boxing (4)
- **Match Reports** (1)
- Sailing and Yachting (4)
- Hockey (4)
- Fencing (4)
- Squash (4)
- Rugby Union (4)
- Tickets Application (4)
- History of FA Cup (4)
- Fitness Equipment (4)

You are approaching the end of your degree course and are interested in pursuing a career as a teacher. Find out about the entry requirements to get on to a post-graduate course in education.

- Finance (4)
- Evaluation (4)
- Library Catalogue (4)
- Optometry (4)
- Archaeology (4)
- **Postgraduate Study** (1)
- Information Services (4)
- Mathematics (4)
- Society and Culture (4)
- Grants (4)
- Languages (4)
- Lectures (4)
- Press Releases (4)
- Residences Office (4)
- Tuition Fees (4)
- Conferences (4)

8.3. Highly relevant target item x not very relevant distractor items

Buy a new scarf to keep you warm on the winter evenings.

Lingerie (5)
Belts (5)
Knit Wear (1)
Clogs (5)
Vintage (4)
Purses (5)
Sunglasses (5)
Football Shirts (5)
Payment Methods (5)
Underwear (5)
Pyjamas (5)
Watches (5)
Shorts (5)
Loafers (5)
Bracelets (5)
Footwear (5)

You are planning a special Christmas dinner with some friends. Find out the price of a group booking for the restaurant at the Hilton hotel.

Gossip (5)
Crafts (5)
Gardening (5)
Photography (5)
Online Trading (5)
Restaurant Search (1)
Awards (5)
Quotes (5)
Television and Radio (5)
Documentaries (5)
Music Guides (5)
Radio Quizzes (5)
Theatres (5)
Celebrities (5)
Trailers (5)
Cosmopolitan (5)

You are struggling with money problems as a student. Find out about what financial aid is available and about applying for income support as a student.

Revision (5)
Research Centres (5)
Eye Clinic (5)
Student Advice Centre (1)
Continuing Education (4)
Courses and Admissions (5)
Postgraduate Enrolment (4)
Health Psychology (5)
Journals (5)
Computer Science (5)
Press Releases (4)
Postgraduate Services (5)
Seminars (5)
Part time Education (5)
Library Catalogue (5)
Clearing (5)

Check the whether forecast for the coming weekend.

Investments (5)
Topical (5)
Public sector (5)
The Economy (5)
Weather (1)
Archived News (5)
Arts (5)
Polls (5)
Travel (5)
Arts Reviews (5)
Films (5)
Motoring (5)
Letters (5)
Current Events (5)
Agriculture (5)
Stock Market (5)

Buy a copy of the Children's book Harry Potter and the Order of the Phoenix for a bit of easy reading.

Psychology (5)
Diaries (4)
Sheet Music (5)
Dictionaries (5)
Art History (5)
Computer Books (5)
Classroom Libraries (4)
Children’s Books (1)
Science Facts (5)
Magazines (5)
National Geographic (5)
True Crime (5)
Popular History4)
Memoirs (5)
Telephone Ordering (5)
Photography (5)

Order a Scented Lily Vase Bouquet for your Mothers birthday.

Funerals (5)
Clay Products (5)
Pens (5)
Cards (4)
Tours (5)
Turning thoughts into Flowers (1)
Posters (5)
Antiques and Antiquities (4)
Christmas Preparation (5)
Gift Vouchers (5)
Sweats (5)
Rings5)
Storm (5)
Sapphire (5)
Rotary (5)
Price Guides (5)

8.4. Moderately relevant target item x moderately relevant distractor items

You are going on holiday. Confirm your flight and check for possible delays.

Group Travel (3)
Guided Tours (4)
Penalty Fares (4)
Locations (3)
Air Fares (4)
Baggage (4)
Speciality Travel (3)
Flight Connections (2)
Travel Accessories (4)
Accommodation (4)
Advanced Booking (2)
Penalty Fares (4)
Boating and Sailing (4)
Sun Holidays (4)
Budget Breaks (4)
UK Breaks (5)

Look for healthy living recipes for a well balanced diet?

Weight Loss (2)
Cooking Accessories (3)
Marie Claire (4)
Spa Treat (3)
Beers and Ciders (4)
Beauty (4)
Basic Cooking (2)
Food Preparation (2)
Fine Wine Centre (4)
Skin Care (4)
Style (4)
Thermal Treatment (3)
Spirits (4)
Bath (4)
Facials (4)
Hair Accessories (4)

Find a seaside hotel for a weekend break.

Guided Tours (3)
Locations (2)
Places to Visit (2)
Holiday Destinations (2)
On a Budget (2)
Surfing (2)
Great Deals (2)
Travel Weather (3)
Outdoor Activity (3)
Sightseeing Tours (3)
Family Activities (3)
United States (4)
Snow Holidays (4)
Security (4)
European Advisor (4)
Baggage (4)

You are planning a weekend away in Bath. Find out the price of a Spa Treat at the new Thermae Bath Spa.

Beauty (2)
Healthy Living (2)
Women's Health (2)
Skin Care (2)
Body Care Treatments (2)
Feng Shui (3)
Consumer Guides (3)
Joint Health (2)
Culture (4)
Recipes (4)
Vitamins and Minerals (3)
Bedding (4)
Marie Claire (4)
Flower Remedies (4)
Body Wraps (4)
Hair Care (4)

You are fan of the television comedy Will and Grace. Buy the first series on DVD.

Entertainment Business (2)
Reviews (2)
DVD Films (2)
Gifts (2)
Soap Quiz (4)
Cinemas (4)
Drama (3)
Charts (3)
Leisure (3)
Film Programs (4)
Song Lyrics (4)
Social Clubs (4)
Quotes (4)
Celebrities (4)
Photography (5)
Documentaries (5)

You are nearing the end of your degree course and are unsure what career to pursue. Find out about graduate recruitment programs and the types of jobs on offer for you.

Media Information (2)
Job Applications (2)
E-Business Strategies (3)
International Business (3)
Capital Markets (3)
Consultancy (3)
Employers Information (2)
Business Profile (3)
Business Directory (3)
Management (3)
Music Industry (3)
Business Cards (3)
Web Marketing (3)
Press Centre (4)
Administration (4)
Web Management (4)

8.5. Moderately relevant target item x not relevant distractor items

Find out the latest on EastEnders and other television soaps.

Hollywood News (4)
Film Festivals (4)
Photography (4)
Galleries (4)
Blue Peter (4)
Advertising (4)
Theatre Memorabilia (4)
Entertainment Business (2)
Video Game Memorabilia (4)
Theatre and Dance (4)
Music (4)
Ticket Sales (4)
Hobbies (4)
DVD Titles (4)
Social Clubs (4)
Events (4)

You are learning to play the guitar, buy sheet music for The Thrills album So Much for the City.

Geography (4)
Book Reviews (4)
Humour (4)
Biographies (4)
Atlases (5)
Thrillers (4)
Book Club (2)
Classics (4)
Signed First Editions (5)
Design (4)
Magazines (4)
Fairy Tales (5)
Textbooks (4)
Comics (4)
Picture Books (5)
Satire (5)

You have just moved to University and are concerned about crime in the local area. For peace of mind, get some home insurance to cover your possessions.

Bank Accounts (4)
Trusts (4)
Cash Payment Scheme (4)
Transactions (4)
Students and Graduates (2)
Debt Consolidation (4)
Security Policy (4)
South Asian Banking (4)
Borrowing Money (4)
Direct Debit (4)
Repaying (4)
Open an Account (4)
Financial Aid (4)
Interest Rates (4)
Accessing Services (4)
Home Loan (4)

Find out about buying tickets for the FA cup final.

Badminton (4)
Sports Equipment (4)
Basketball (4)
Football News (2)
Wind Surfing (4)
Rugby Union (4)
Swimming (4)
Mobile Text Alerts (4)
Snow Boarding (4)
Boxing (4)
Opinion and Commentary (4)
Shooting (4)
Cricket (4)
Sailing and Yachting (4)
Hockey (4)
Sports Village (4)

You are over sleeping for too many lectures. Buy a radio alarm clock to wake you up in the mornings.

Communication Devices (4)
Toasters (4)
Radios (2)
CD-ROMs (4)
Computer Security (4)
Cameras (4)
Multimedia (4)
Operating Systems (4)
Cables (4)
Services (4)
Minidisc Player (4)
Product service (4)
Car Audio (4)
DVD Portables (4)
Keyboards (4)
Music Technology (4)

You are interested in doing work related to your course this summer. Find out about a summer placement job at the computer company IBM.

Survey Reports (4)
Business Profile (4)
Legal Issues (4)
Website Design (4)
Transport and distribution (4)
Human Resources (4)
Careers Information (2)
Presiding Officers (4)
Administration (4)
Office Essentials (4)
Press Centre (4)
Stationery (4)
Marketing (4)
Employment Data (4)
Business Booking (4)
Health and Safety (4)

8.6. Moderately relevant target item x not very relevant distractor items

You are planning a party with some friends. Order a case of fine wine for the party.

Skin Care (5)
Astrology (5)
Pottery and China (5)
Sauna (5)
Cutlery (5)
Antioxidants (5)
Bathroom (5)
Beers and Ciders (4)
Carpets (5)
Detoxification (5)
Herbal Remedies (5)
Joint Health (5)
Parenting (5)
Yoga (5)
Vitamins and Minerals (5)
Fitness (5)

You are going to visit friends in London at the weekend. Find out about current problems and closures on the London underground.

Accommodation (5)
Hospitality (5)
Confirm Flight (5)
Mountain Biking (5)
Snow Holidays (5)
Surfing (5)
City Breaks (2)
Air Fares (5)
Airport Services (5)
Boat Hire (5)
Canada (5)
Charter Flights (5)
Hawaii (5)
Lodging (5)
Sun Holidays (5)
Timeshares for Sale (5)

You have recently received the reading list for your Psychology degree course. Try and order the required textbooks for the course.

Geography (5)
Maps (5)
Horror (5)
Fiction (5)
Theory Books (2)
Rocks and Fossils (5)
Photography (5)
Popular Biography (5)
Biographies (5)
Romance and Love (5)
Historic Figures (5)
Travel Guides (5)
Magazines (5)
Thrillers (5)
Telephone Ordering (5)
Calendars (5)

Set up a direct debit payment scheme to make regular payments towards your television licence bill.

Car Insurance (5)
Change of Address (5)
Complaints (5)
Current Accounts (3)
Postal Prices (5)
Legal Glossary (4)
Parcel Services (5)
Borrowing Money (5)
Health Insurance (5)
Pensions for Private Clients (5)
Repaying (5)
Offshore Banking (5)
Personal Finance (5)
Tax-Free Savings (5)
Motor Insurance (5)
Card Protection (5)

It is your nephew’s birthday and he would like a Rocking Horse. Found out how much money a Rocking Horse is likely to cost.

Science Fiction (5)
Television Games (5)
Classic Toys (2)
Puzzles (5)
Karaoke Systems (5)
Competitions (4)
Music Games (5)
Action Figures (5)
James Bond (5)
Radio Control (5)
Internet Games (5)
Building Toys (5)
Sticker Print Packs (5)
Brain Games (5)
Cartoons (5)
Croquet (5)

Check your bank balance

Loans (5)
Credit Cards (4)
Mortgages (5)
Small Business Briefing (5)
Travel Services (5)
Investment (5)
Customer Service (2)
Payment Solutions (3)
Marketing Guide (5)
Insurance (5)
Assessing Services (4)
International Trade (5)
Tools (5)
Publications (5)
Pensions (5)
Franchising (5)

8.7. Filler trials

Find out when the summer ball has been scheduled for?

Summary and Results
Tuition Fees
Term Dates
Publications
Policies
University Research Centres
About Research
Accommodation
Health and Welfare
Money Issues
University Events
Lectures and Seminars
Discussions and Reading Groups
Concerts and Recitals
Exhibitions
Conferences

Check and see if your recorded parcel arrived safely at its destination.

Mail Services at Home
Delivery Services
Parcel Services
Special Deliveries
Business Mail Services
Moving Home
Save Money on Fuel Bills
Stamps
Address Management
Travel Services
Money and Banking
Phone Services
Greeting Cards
International Services
Postal Prices
Licenses

Find out about becoming a student volunteer?

Advice
Contact
Careers
Disabled
Mental Health
Homelessness
Learning Disabilities
Autism
Guidelines
Location
Latest News
Children
Roles
How to Get Involved
Mission Statement
Elderly

Read reviews on recent computer game releases.

Performance and Capacity
Artificial Intelligence
Human-Computer Interaction
Product Support
Operating Systems
Newsgroups
Education
Cyber Cafés
Games
Online Learning
Jobs
Internet Marketing
Publications
Bulletin Board
Free Web Space
Radio Stations