In Proceedings of the Sixth International Conference on Cognitive Modeling, 46-51.
Mahwah, NJ: Lawrence Earlbaum.

Good Enough But I'll Just Check: Web-page Search as Attentional Refocusing

Duncan P. Brumby (BrumbyDP@Cardiff.ac.uk)
Andrew Howes (HowesA@Cardiff.ac.uk)
School of Psychology, Cardiff University,
Cardiff CF10 3YG, Wales, UK

Abstract

When people search a set of Web labels for links that are relevant to their information goal, they attend to the labels and estimate the likelihood that the link will lead to the goal. Recent findings indicate that people sometimes, but not always, assess only a subset of the links available. We report an ACT-R model of Web-page search that was inspired by Young's (1998) rational account of exploratory choice, but which was also sensitive to the psychological constraints encoded in the ACT-R theory of the human cognitive architecture. The behavior of the model differs substantially from previous ACT-R models of Web navigation, at least for single-page search. We describe an experimental test of the models behavior and qualitative and quantitative fits of the model to the data.

Introduction

Consider searching a newly encountered Web page for links that are relevant to the achievement of some search goal. During this activity, people focus on the labeled links in order to derive a subjective assessment of the likelihood that the selection of a given link will lead to the achievement of the current information goal. It is known that people tend to select links to Web sites with labels that are more relevant to their current goal (Pirolli & Fu, 2003). In a simplified single page menu search task, however, Brumby and Howes (2003) observed that people do not always assess all of the items available in the choice set prior to the selection of an item, and that they re-fixate a smaller and smaller subset of these items prior to selection. In the current paper, we present a model of menu search behavior developed with the ACT-R framework (Anderson & Lebiere, 1998). The model makes use of ACT-R's architectural assumptions in order to produce behavior that (1) provides both qualitative and quantitative fits across a range of performance measures including eye-movement data, and (2) is broadly consistent with a rational analysis of exploratory choice (Young, 1998). We also report an experimental test of the model.

In a simplified single-page Web search task, Brumby and Howes (2003) asked participants to search single-level Web pages in pursuit of an information goal. The pages contained a single goal, or *target*, link and the rest of the links were *distracters*. An eye tracker recorded participant's eye movements while they searched the labeled links. The relevance of the distracter links to the goal was varied while the quality of the target link was held constant.

As we have said, Brumby and Howes found that participants did not always fixate all of the links available in the choice set prior to selection, and that they re-fixated a smaller and smaller subset of these links prior to selection. These signature behaviors are consistent with qualitative observations of people freely exploring a novel device interface (Rieman, 1994) and with the search of early keyboard-driven menus in which only a single choice can be seen at a time (MacGregor, Lee, & Lam, 1986).

Brumby and Howes found that participants selected the target link immediately after its first fixation, i.e. they satisficed, on 31% of trials and repeatedly rescanned a smaller subset of the links prior to selection on 69% of trials. They also found that participants sometimes fixated spatially adjacent links consecutively in time and sometimes skipped links, as observed by Byrne (2001) and Hornof (in press). Interestingly, participants more frequently skipped links after they had fixated the item that they would eventually select.

Further, Brumby and Howes found that the presence of lower relevance distracters resulted in fewer items being fixated. This was because participants who were given lower relevance distracters, and target links of the same relevance, were more likely to satisfice, i.e. to select a target link immediately after fixating it for the first time.

While it is known that the relationship between eye-movements and cognitive processing, in this case assessment, can be complex, we assume that the fact that Brumby and Howes' participants were more likely to satisfice if the distracters were of lower quality is evidence that participants made assessments of fewer links. This finding therefore supports the idea that people may adjust an independent assessment of the relevance of a link, in order to derive an estimate that is interdependent with the quality of the other links in the choice set. Consequently, 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.

A number of models of Web navigation and menu search have previously been reported, including models constructed in ACT-R (Byrne, 2001; Salvucci, 2001) and also in EPIC (Hornof, in press; Kieras & Meyer, 1997). These models have focused on the basic perceptual/motor components involved in simple, routine menu search and are not able to model the semantically laden assessment process that characterizes complex menu search, typical of Web navigation.

SNIF-ACT (Pirolli & Fu, 2003) makes an important contribution to the development of models that can simulate users searching the World Wide Web for information relevant to an unfamiliar information goal. The model provides a high-level characterization of Web search behavior, focusing on link selection and site leaving behaviors. Importantly, the model uses ACT-R's declarative memory module in order to derive assessments of a menu labels relevance to the information goal (what Pirolli & Card, 1999, term *information scent*). The model accurately predicts that users will select menu labels that have high information scent. However, SNIF-ACT lacks a plausible model of how people search an individual page. This over simplification is non-trivial, if people do not always assess all items on a menu page prior to selection.

While some previous models do not capture the menu search behavior observed by Brumby and Howes (2003), Young (1998), and more recently Cox and Young (submitted), have presented a rational analysis of exploratory choice, which provides an important theory from which to understand people's menu search behavior. These models extend Anderson's (1990) rational analysis of problem solving to the uncertain environment inherent in menu search. In these models, the underlying choice at each cycle is between selecting the item judged most likely to lead to goal immediately, or performing another (re)assessment. The decision between action and assessment is based upon a simple utility function. In other words, items are assessed so long as the expected gain of making another assessment exceeds the cost of the assessment.

Young's (1998) model is particularly interesting because it is sensitive to the implications of the structure of the task environment in many menu search tasks. When searching a menu that contains a given choice set of menu items ($item_1$ … $item_n$), typically only a single item will lead to the achievement of the information goal. Given some probability estimate of this likelihood it can be assumed that the sum of estimates across all items in the choice set must be equal to one. 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, p. 474). A novel prediction to emerge from the normalization assumption is that the relevance of both the target *and* the distracters will affect the decision of whether to select or continue assessment.

Web-page search as attentional refocusing

We present an ACT-R model of our previous menu search data (Brumby & Howes, 2003). The model is partially constrained by Young's (1998) rational analysis of exploratory choice, and in addition is consistent with the memory constraints imposed by ACT-R (Anderson & Lebiere, 1998). Our aim was that the model should at least demonstrate the observed search behaviors: Participants typically do not fixate all of the items in the choice set, and often items are fixated on multiple passes prior to selection.

In order to capture these behaviors it is necessary to model how people choose between selection of an item and further assessment of items. A candidate mechanism is ACT-R's conflict resolution mechanism. Competition between productions for control of behavior is governed by a mechanism derived from Anderson's rational analysis of choice. However, as Young and Cox (2000) point out, although the production rule system in ACT-R takes account of past experience in operator selection (Lovett & Anderson, 1996), the parameters associated with a given production rule change relatively slowly in order to reflect the long-term learning of production utility. Consequently, our approach to this problem was to further explore ACT-R's declarative memory retrieval mechanism. This approach is consistent with that employed in previous ACT-R models of Web navigation (Pirolli & Fu, 2003) but makes a novel use of the mechanism by which source-activation models the focus of attention.

The key features of the model were:

1. The ACT-R goal chunk included *n* slots, one for each label link and each of which, initially, had a value of *unassessed*. We call these *assessment* slots. The goal also has two additional slots, one for the current information goal and the other for the current attended visual location.

2. Assessment of label links was achieved by repeated attempts to retrieve chunks from declarative memory (Figure 1). Chunks represent world knowledge. The successful retrieval of a chunk was assumed to indicate that there was positive information linking the label and the goal, and resulted in the replacement of *unassessed* values on the goal with the retrieved value. An unsuccessful retrieval resulted in the replacement of a slot value with "not relevant", i.e. equivalent to setting the slot value to nil.

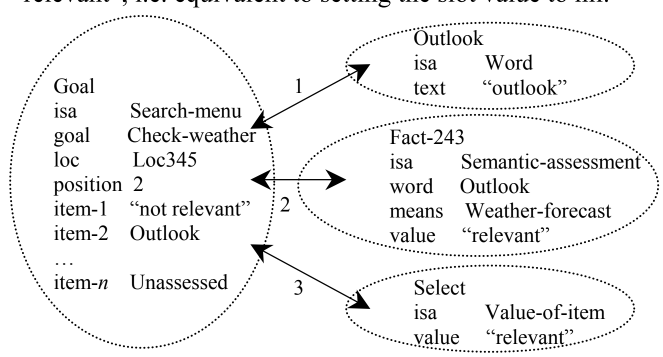

Figure 1: An example of assessment chunks for a menu item. The goal chunk contains slots for $item_1$ … $item_n$ that can take the values *Unassessed*, *Word-n*, or *"not relevant"*. An iconic representation of the currently attended menu item is passed from the visual buffer (*"outlook"*) to signal a retrieval request for a syntactic assessment chunk (arrow 1), which if retrieved leads to requests for further assessment chunks to be retrieved (e.g. arrow 2 → arrow 3). Decision to select the item is signaled if all assessment chunks are retrieved. Attention is shifted to another menu item if there is a failure to retrieve any of the assessment chunks.

3. Whether or not a chunk was retrieved depended in part on its activation. In ACT-R the activation is determined by a combination of base-level activation B, spreading activation $W_j \cdot S_{ji}$, and a transient noise.

$$A_i = B_i + \sum W_j \cdot S_{ji} + \varepsilon \qquad (1)$$

The former was set to zero in our model and played no role in capturing the desired behaviors. Spreading activation played a crucial role (see bullets 4 and 5 below).

4. The amount of spreading activation received by a chunk depended in part on the attentional focus. The more label links that were in the choice set (as represented by the assessment slots) the lower the amount of source activation W received by the chunk. Conversely, the fewer the number of items in the choice set the greater the amount of source activation W received by the chunk.

$$W_j = 1 / (n + 2) \qquad (2)$$

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

$$S_{ji} = C + M \cdot \sigma(j, i) \qquad (3)$$

Where, the σ(j, i) component reflects the input measure of semantic similarity between chunks *j* and *i*, and C is a negative quantity that serves as a base of associative strength and M is a positive multiplier. Moreover, 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 (see Table 1).

6. Different chunk-types were used to represent different types of assessment (e.g. syntactic or semantic). Importantly, the model could sometimes retrieve some of the assessment chunks for an item, while other assessment chunks for the item would fail to be retrieved (see Figure 1). In this case the item was judged partially relevant to the information goal. Given some change in the value of source activation W, future assessment of the item may warrant selection. Thus, the model maintained pointers on the goal chunk to such partially relevant menu items in order to allow the item to be reassessed.

7. The model (re)assessed another item 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 an unassessed item (nearest the current visual location) and reassessing a previously relevant item (see bullet 6 above), compete in a stochastic selection process. The utility of the production rule for assessing an unassessed item was greater than the utility of the production for reassessing an item.

8. The model chose to select an item if it had been judged to be highly relevant to the current information goal, indicated by retrieval of all chunk-types related to the item.

In summary, if the assessed distracter links were not relevant to the current information goal, then the model was more likely to make a positive assessment of a labeled link, and furthermore was more likely to select a highly relevant link immediately. This behavior is consistent with Young's (1998) normalization assumption and reflects a constraint imposed by the structure of the task environment.

Menu Search Experiment

The models performance was compared with data collected from a menu search experiment that extended our previous work (Brumby & Howes, 2003). In the study participants searched single-page menus. All menus contained a single goal item and the participant's task was to select the goal item as quickly and accurately as possible. Participants did not progress to the next trial until they had selected the goal item. Independent ratings were taken of label quality. The menus differed in terms of the quality of the goal item (very good or moderately good) and the quality of the distracter items (moderate, bad, or very bad). The quality of a label was determined from participant's ratings of the degree to which an item was relevant to the achievement of the search goal. Participants (n = 20) were native English speakers. Eye tracking was performed using an ASL Pan/Tilt optic eye-tracking system.

Results and Discussion

It was found that the choice between continued assessment and selection in menu search was affected by the quality of both the goal item that was selected and quality of the assessed distracter items. Figure 2 demonstrates that (1) participants rarely fixate all of the items available in the choice set prior to the selection of an item, and (2) they re-fixated a smaller and smaller subset of these items prior to the selection of an item. In particular, when the quality of the goal item was held constant, participants fixated fewer of the items fewer times prior to selection of an item when the quality of the distracter items that were assessed were less attractive ($F(2, 38) = 3.66$, $p<.05$). Indeed, if we

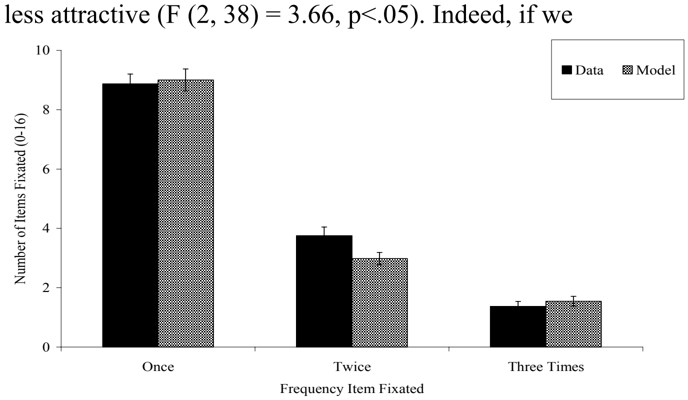

Figure 2: Data and model fits for the number of items fixated at least once, twice, and three times (collapsed across condition)

consider only the number of items fixated at least once, then this effect amounted to participants on average fixating two items (approx. 12.5%) fewer when the distracters were very bad in quality compared to moderate in quality. This key finding provides evidence of interdependence in menu choice assessment: whether people even bother to fixate some items depends on the quality of the distracters that they have assessed. There was also significant effects of quality of distracter item on time to selection ($F(2, 38 = 8.71$, $p<.001$), accuracy of selection ($F(2, 38) = 107.54$, $p<.001$), the number of items fixated after the initial fixation of the goal item ($F(2, 38) = 7.75$, $p<.005$), and a significant trend for the proportion of trials in which participants selected the goal item immediately after fixating it for the first time ($F(2, 38) = 3.18$, $p<.1$).

Furthermore, we found that when the selected goal item was more relevant to the current information goal, participants were quicker to select that item ($F(1, 19) = 258.24$, $p<.001$) and were more accurate with that selection ($F(1, 19) = 174.63$, $p<.001$), and fixated fewer items after the initial fixation of the selected item ($F(1, 19) = 84.37$, $p<.001$). Indeed participants were more likely to select the goal item immediately after fixating it for the first time ($F(1, 19) = 46.69$, $p<.001$). Although intuitively obvious, it is not clear that previous models of menu search would predict some of these findings, e.g. SNIF-ACT (Pirolli & Fu, 2003).

Model Fitting

The primary aim in evaluating the validity of the model was to match the models performance across a range of dependent variables used in the menu search experiment. Given that the primary focus of our study was on eye movement data, we made the assumption that ACT-R's movements of visual attention to items in the menu can be taken to broadly match a participants eye movement fixation centered around a menu item in our menu search task. To this end, the model used Salvucci's (2001) EMMA system to provide a more detailed theory of visual encoding. The ACT-R models output was therefore subject to the same analysis as a participant's eye movement fixation sequence.

The ACT-R model interacted with a menu that was the same as that searched by participants in the experiment. Although, we gathered participant ratings of the degree to which menu labels from the experiment were relevant to the achievement of the search goal. It was not possible to use these ratings as input to the model. Participant's ratings were provided on a discrete scale, whereas the model required a probability-like quantity (see bullet 5). Furthermore, as described in bullet 6, it was assumed that labeled links were evaluated in stages, involving several assessments of the label (Young, 1998). The input similarity values were therefore estimated for labels of different qualities and were essentially free parameters in the model (see Table 1 for the estimated values). In Table 1 the label quality 'moderate distracter', for example applied to all distracter labels in the experimental condition where the distracters were of moderate relevance to the search goal.

Table 1: Estimated Similarity Values

	Chunk-Type		
Label Quality	Word	Semantic-Assessment	Value-of-Item
Very Good Goal	.80	.45	.25
Moderate Goal	.50	.20	.18
Moderate Distracter	.25	.15	.10
Very Bad Distracter	.20	.10	.05

In fitting the model to the data, we systematically attempted to maximize the fit of the models performance across three of the four[1] experimental conditions on the main dependent variables (number of items fixated at least once, twice, and three times; accuracy of selection; percentage of self-terminating searches; time to selection; percentage of items fixated after the initial fixation of the goal item).

In order to estimate the input similarity values for each label of the same quality (see Table 1), we went through each experimental condition in a piece meal fashion, and obtained input values that maximized the models fit, over 100 model runs, across the range of dependent variables. Once an input similarity value was estimated, we 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 *very good goal, very bad distracter items* condition, input similarity values for the chunks associated with very good goal chunks and very bad distracter items chunks were estimated which provided a good fit with the data. Next, fitting the model to the *moderately good goal, very bad distracter items* condition, we held constant the estimated input similarity values for the chunks associated with very bad distracter items, and varied the value of the chunks associated with the goal item only.

Importantly, all of the models free parameters were estimated from fitting the model to three of the four experimental conditions. Thus, the models performance on the final condition (*very good goal, moderate distracter items*) was predictive (in the sense that the model was not iteratively fitted to the data).

The model provided a good fit with the data across most of the dependent variables. The model demonstrated the observed signature behaviors: (1) the model rarely attended all of the items available in the choice set prior to the selection of an item, and (2) re-attended a smaller and

[1] The actual design included six experimental conditions, however, post-hoc analysis revealed non-significant pairwise differences between very bad quality and bad quality distracter items across quality of goal. For simplicity we focused on the contrast between moderate and very bad quality distracters across very good quality goal and moderately good quality goal (i.e. we did not model the two bad quality distracter conditions).

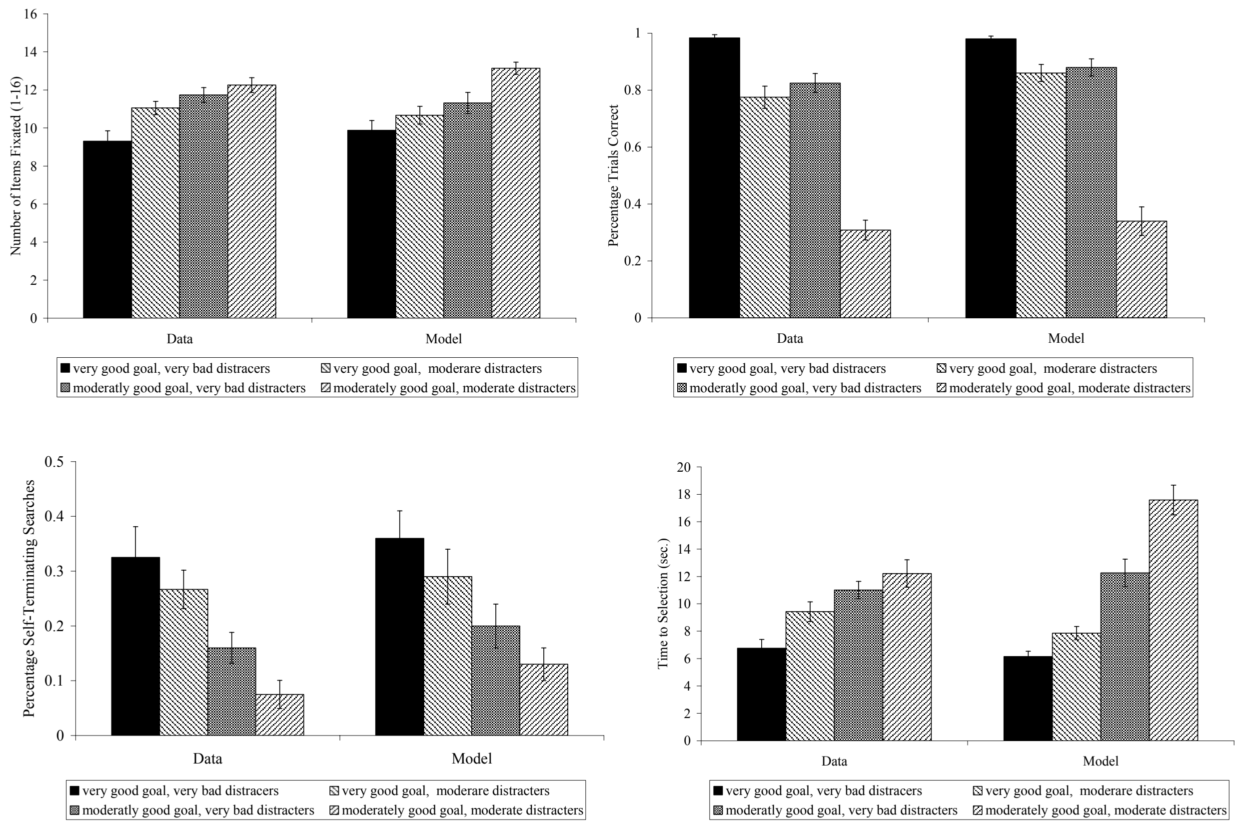

Figure 3: Data and model fits for effect of quality of goal and quality of distracter items on the (1) number of items fixated; (2) percentage of trials correct; (3) percentage of self-terminating searches; and (4) time to selection

smaller subset of these items prior to the selection of an item (see Figure 2 for aggregate model fit). Across conditions, the model was able to account for shifts in this behavior (r^2 = .96), i.e. attending fewer items when the distracter items were very bad in quality. This correlation was highly significant ($F(1, 10) = 214.08$, $p<.001$). In particular, Figure 3 shows that the model provided very good fits to the data for (1) the number of items fixated; (2) percentage of trials correct; (3) percentage of self-terminating searches; and (4) time to selection. The model did not provide a good fit for the number of items fixated after an initial fixation of the goal item, however.

General Discussion

We have presented a model of how participants searched single simplified Web pages in order to find links that were relevant to an information goal. In contrast to previous models of Web navigation (e.g. SNIF-ACT, Pirolli & Fu, 2003), but consistent with the data, the reported model sometimes selected a link before evaluating all links. The model was built in ACT-R and took advantage of the fact that in the architecture a fixed amount of source activation is distributed among the declarative chunks that are associated with the goal. The search process was therefore modeled as attentional refocusing. The probability of retrieving knowledge that associated a label with the goal was therefore dependent on the number of other labels in the attended choice set. In other words, the estimated likelihood that any single label would lead to the desired information was partly dependent on the values of other items. This interdependency reflects Young's (1998) normalization assumption.

The model was evaluated by comparing its predictions to a range of measures. In an experiment, the relevance of the goal item and the distracter items to the information goal was manipulated. The data were consistent with the prediction that people would fail to assess all of the items in the choice set, and that they would repeatedly reassess a smaller and smaller subset of potential candidate items prior to selection.

The model was inconsistent with the data in at least one respect. It did not provide a good fit to the number of items fixated after the initial fixation of the eventually selected item. This discrepancy may be partially explained by the systematic search strategy used in the model. Consistent with previous models of routine menu search (e.g. Byrne, 2001; Salvucci, 2001), the model attended each item in turn in the menu in a top-to-bottom fashion, with the exception of choosing to reassess an item. Our analysis of participant's eye movement fixation sequences revealed that the distance between contiguous fixations was typically in the region of

1.5 items, suggesting participants regularly skip over items. Our model does not capture this item-skipping strategy.

Hornof (in press) proposed a maximally efficient foveal sweep strategy, in which multiple items within the fovea (defined as one degree visual angle) are assessed with a single fixation. Interestingly, in our data the foveal sweep strategy (for fixating previously unfixated items) was moderated by whether or not the participants had already fixated an item that was attractive. However, it is not clear that participants in our experiment were assessing multiple items with a single fixation, because items were approximately one-degree visual angle in height and the distance between items was also approximately one-degree visual angle. Further theoretical work is required.

The model does not capture all of our previous data (Brumby & Howes, 2003). In particular, we have found that the history of information search moderates the local search strategy. After participants had completed trials in which they were more likely to select an incorrect item (because the distracters had been made more attractive), they were more cautious about selection. That is, they assessed more of the items in the choice set and were less likely to select an item immediately following an initial fixation of that item. Extending the current model to account for these findings should be relatively straight forward, because ACT-R's production rule learning mechanism is well suited to modeling the influence of history of successes on operator selection (Lovett & Anderson, 1996).

The idea that Web-page search is attentional focusing may seem counter-intuitive. In our model the goal of assessing unassessed items reduces the probability of retrieving information about the currently fixated item. Although, this mechanism predicts the observed behavior 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. Further data is required.

In summary, the aim of the work reported in this paper was to build a model of single-page Web-search that was constrained by ACT-R architectural assumptions and by Young's (1998) rational analysis of exploratory choice. The model was supported by the results of an experiment that tested the consequences of distracter-quality and goal-quality for eye fixations during the search of simplified Web pages.

Acknowledgments

This work was partially funded by an EPSRC studentship to Duncan Brumby. An earlier version of this work was presented at the 1st UK ACT-R Learning Group held at UCL, London UK, January 2004. We thank Richard Young and Anna Cox for comments on this work, and three anonymous reviewers.

References

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

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

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

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

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

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

Hornof, A.J. (in press). Cognitive strategies for the visual search of hierarchical computer displays. To appear in *Human-Computer Interaction*

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

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

MacGregor, J., Lee, E., & Lam, N. (1986). Optimizing the structure of database menu indexes: a decision model of menu search. *Human Factors, 28(4),* 387–399

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

Pirolli, P., & Fu, W-T.F. (2003). SNIF-ACT: a model of information foraging on the world wide web. In *Proceedings of the Ninth International Conference on User Modeling,* 2003

Rieman, J. (1994). *Learning strategies and exploratory behaviour of interactive computer users.* PhD dissertation, University of Colorado, Boulder, CL

Salvucci, D.D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1(4),* 201– 220

Young, R.M. (1998). Rational Analysis of exploratory choice. In M.Oaksford & N.Chater (Eds.). *Rational Models of Cognition.* Oxford: 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 the proceeding of the *Seventh Annual ACT-R Workshop and Summer School,* Pittsburgh, PA, 2000