In Proceedings of the Sixth International Conference on Cognitive Modeling, 82-87.
Mahwah, NJ: Lawrence Earlbaum.

A Rational Model of the Effect of Information Scent on the Exploration of Menus

Anna L Cox[1] (a.cox@ucl.ac.uk) & Richard M Young[2] (r.m.young@herts.ac.uk)
[1]UCLIC, Dept of Psychology, University College London, London, UK
[2]Dept of Psychology, University of Hertfordshire, Hatfield, UK

Abstract

This paper describes a computational model based on rational analysis which is compared to experimental data and enables us to explain the number and order of assessments that participants make on the items on a web-page. The model provides both a good fit to the scan paths of the participants and an explanation for why a particular item is chosen. The strength of this research is that it provides a qualitative account of why particular patterns of scent result in particular scan patterns. The research provides further support for the rational framework as an explanation of exploratory behaviour.

Introduction

In the real world people often do not read the instructions for a novel system either through choice, or because they are not available. Instead, they learn how to use the novel system by actually trying to use it, drawing on a combination of prior knowledge, information from the interface itself such as the semantic similarity between the labels and their task description, and problem solving skills. This phenomenon is described as exploratory learning.

When exploring an interface, people might select a particular option for one of two reasons: because they believe that it will lead them to the goal (focused exploration); or because they believe that it will provide them with some information about the system they are interacting with (free exploration) (Cox 2002).

Our earlier research (Cox 2002) led to the development of a rational framework for modelling this kind of exploratory learning. We have already used it to explain empirical findings such as why people do and learn different things from free as against focused exploration (Young & Cox, 2000). We are now using it as a framework to build models of single-level menu exploration which we compare to data gained from empirical studies.

The framework and how it works

To begin, we will consider the behaviour of a model built within the framework when engaged in free exploration. At any moment, the model has a number of possible things it can do. These might be clicking on a hyperlink, considering some feedback that has just been received, reading a label, formulating a hypothesis to test, etc. These are referred to as Exploratory Acts or EAs. EAs differ in the number of stages they consist of and therefore in the amount of time and effort required to carry them out. For example, pressing a button or clicking on a hyperlink would require a single quantum of effort, whereas interpreting feedback from the interface might require a number of quanta: one to read the display; another to realise that the information has changed; and a third to understand the type of change.

Each EA has a cost (C) and a value (ΔI) determined by the costs and values of its constituent quanta. Each quantum has a cost reflecting the time it would take to perform it, and a value related to the estimated increase in information that would be gained. The cost of the EA that proposes a single button press therefore is likely to be low as it only requires one quantum of processing, but the value may be high or low depending on what is known about the button already. The cost of the EA that proposes interpreting some feedback would be higher than that of the button press as this EA requires more quanta of processing (three in this example) and again the value would vary depending on the amount of information expected to be gained.

We assume that which of the proposed EAs is chosen in any given situation is determined by rational analysis. This theory suggests that when trying to learn about the device, the only reason to choose an EA will be the one that is believed to elicit the highest amount of information (ΔI) for the least cost (C): $\Delta I/C$.

During focused exploration, the model has a specific goal to achieve and therefore an EA could be chosen for one of two reasons. Firstly, in the same way as for free exploration, an EA might be chosen because it will elicit information that is not known, or secondly, it might be chosen if it is expected to lead to the goal. In this situation, therefore, the efficiency of some EAs is calculated as the expected amount of information elicited about the device per unit of cost, and for some others the efficiencies are calculated as the probability (P) that they will lead to the completion of the goal (G) minus the cost (C) of getting to the goal (PG-C). This results in the behaviour suggested by Pirolli & Card's scent following theory (1999).

The framework describes a cycle of three stages. In stage one, the efficiency of all the EAs possible at that moment are calculated. The efficiency of an EA is equal to the expected amount of information gained by the EA, divided by the cost of executing it. In stage two, the EA with the highest efficiency is chosen. The model will therefore choose whichever EA proposes the highest information gain (ΔI) per unit of cost (C) or the greatest probability of reaching to goal (PG) minus the cost (C). Finally, in stage three, the chosen EA is executed.

Menu Search Data

In response to an earlier model of menu search (Young, 1998), Brumby and Howes (2003, 2004) set out to test the hypothesis that people would assess fewer poor distracter items than mediocre distracter items. Participants were shown a single page on which a series of hyperlinks were arranged in a vertical list and asked to identify which of the menu items they would choose in order to find the goal information. Eye-tracking was performed using an ASL Pan/Tilt optic eye-tracking system.

Results

Average performance across menus. Rieman (1994) noted that users often appear not to assess all possible items before making their selection. Initial analysis of the empirical data supported this observation and suggested that on average, participants looked at about half the items in the menu before making their selection (Brumby & Howes 2003).

Performance on Individual Menus. More detailed analysis of the traces provided by the eye-tracker shows that in fact there are two kinds of behaviour that can be observed. The first is where a participant appears to continue scanning the menu in search of the goal item after fixating on the item eventually selected. This was the most common behaviour and occurred on 69% of trials. The second is the opposite of this, when the participant selects the goal item immediately after fixating on it (self-terminating behaviour). This occurred on 31% of the trials.

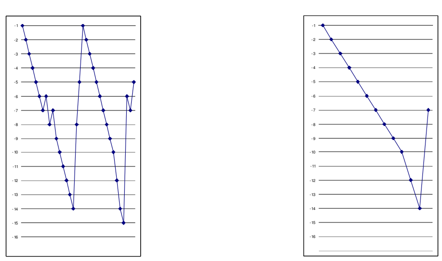

Figure 1: Illustrations of continued scanning of a menu.

The illustrations in figures 1 and 2 show a simplified version of the scan paths of participants in the experiment. The menu items are arranged in order downwards on the vertical axis, while time runs to the right. The scan paths show which items were fixated and in what order.

The first illustration in figure 1 shows an example of continued scanning of the menus. This pattern was seen in approximately 32% of the continued scan cases (22% of the total). The second illustrates an example where the participant continues to fixate on the menu items below the goal item before returning to that item and selecting it. This pattern was seen in approximately 68% of the continued scan cases (47% of the total).

Although the eye-tracking data suggests that sometimes the participants do not fixate on all the items in the list, but, as in figure 1, stop short of the end of the menu, we suggest that this is not sufficient evidence to suggest that the last two items in the list are not being assessed. Byrne et al (1999) provide evidence from eye-tracking studies of visual search in which participants tend to start at, or near, the top of a list and search downwards sometimes skipping items. This behaviour is explained in Salvucci's (2001) model of menu search in which he noted that although the model did not fixate on some items, it did attend to and encode them.

Figure 2 illustrates the other scan pattern from the eye-tracking data: the participant assesses each of the items in the menu until he locates the goal item, and then selects it immediately, without fixating on any other items in the menu.

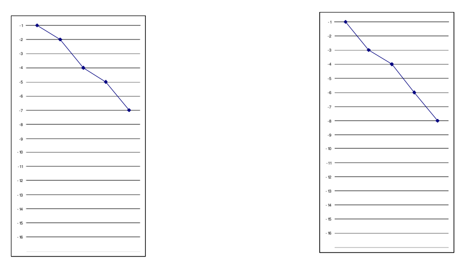

Figure 2: Illustrations of immediate selection following fixation on the goal item (self terminating behaviour).

The Model

In this paper we report the behaviour of a model built within the framework described above. Previous research suggests it is not straightforward to implement the framework in ACT-R (Young & Cox 2000) and therefore the framework has been built in Lisp. It is a model of the cognitive processes involved in an agent's *assessment* of a novel menu presented on a web-page. The agent has the goal of selecting the item that will lead to goal completion and therefore this is focused exploration. However, as the menu presented is novel, the first thing that the model has to do is to gain some information about the menu that will then enable it to conduct focused exploration (i.e. use a scent *following* strategy such as PG-C). The model presented here stops short of the scent *following* part of the task (i.e. selecting the item that will lead to the goal) and as a consequence it is not necessary for this model to make use of the two measures of efficiency. Instead, this model incorporates only the ΔI/C efficiency measure. In order to extend the model to include scent *following*, an additional type of EA (SELECT) would need to be incorporated. The efficiency value for this EA would need to be calculated using PG-C rather than ΔI/C.

The model interacts with a representation of a single level menu that consists of sixteen items as used by Brumby & Howes (2003). Each menu is associated with a particular goal: e.g. Find out who stars in the film "The Lord of the Rings".

The model includes two types of EA: 'assess information SCENT' and 'ANTICIPATE the result of selecting this item'. The SCENT EA should be thought of as being an amalgamation of perceiving the label, reading the label (at a lexical level) and considering the semantic similarity between the label and the current task. The ANTICIPATE EA should be thought of as some additional cognitive effort that considers whether the label is likely to lead to the goal. For example, given the goal of finding an armchair on a furniture website, the ANTICIPATE EA models the agent's consideration of whether the label 'home' is likely to lead to the homepage of the site, or to a list of home furnishings. Each of these EA types has a cost associated with it with the ANTICIPATE EA type being more expensive in mental effort than the first type. There is also a fixed cost of moving attention from one item in the menu to the next.

Brumby and Howes (2003) asked participants to rate the relevance of each item to the goal, on a five point scale. We consider the median ratings from the participants to be a good indication of the information scent of each menu item to the relevant goal. These ratings have been used as the judgments made by the model regarding the amount of information scent from a particular item. These are used to represent the value of the outcome of the SCENT EA in the model.

The ratings provided by participants were also used as the basis upon which to calculate the conditional probabilities of an item leading to the goal or not, given each scent rating. Before assessing any items, the model 'knows' the number of items in the menu. Each of these items is considered to be equally (ir)relevant to completing the task. The scent ratings provided by the participants are also used as the basis for determining the new relevance (R) value of an item following an assessment. On each page, the set of relevances R_i are mapped into a set of probabilities P_i by the transformation $P_i = \text{odds}(Ri)/\sum\text{odds}(R_j)$, where odds(R) is defined in the standard way as odds(R) = R/(1–R). Note that $\sum P_i = 1$, reflecting the fact that exactly one option on the page leads to the goal.

Results

The model was run on the same menus that were used in the empirical data collection (Brumby & Howes 2003). The goal item in each menu had a high scent value (1) and appeared in various positions within the menus. The menus varied in terms of the quality of the distracter items. They were either low quality (all distracters had a scent value of 5) or mediocre quality (mean scent value of the distracter items was 3). The output trace from the model identifies which EA fired on each of the cycles. Average performance of the model across menus is briefly considered together with performance on individual menus. The performance of the model is compared to that of human participants.

Fitting the model. No substantial effort has been made to manipulate the parameters of the model in order to ensure a good fit between the quantitative data from the empirical study and that from the model. However, even without this, the model appears to perform similarly to Brumby & Howes' participants and their ACT-R model (Brumby & Howes 2004) in so far as exhibiting two tendencies:

I. The model tends not to assess all of the items in the choice set.
II. Items are often assessed on multiple passes before selection.

Looking more closely at the model runs on individual menus shows that the model behaves similarly to the participants. The following sections describe the model's behaviour on the menus in detail.

Good goal, poor distracters. Just over half (8/14) of the menus used in Brumby and Howes' (2003) experiment were made up of a good goal (scent value 1) and 15 poor distracters (scent value 5). These eight menus can be thought of as being identical to each other except for the position of the high scent item. Of these menus, two were completely identical as far as the model was concerned as the high scent item was in the same position in each: position 1. On these menus the model chooses to do a SCENT assessment on the first item, and then does the ANTICIPATE assessment on that item immediately and then stops. (If scent *following* were also to be implemented in the model, the model would choose to SELECT this item at this point.) In concordance with the eye-scanning data, the model predicts that this pattern of behaviour is most likely to occur on menus where there is (at least) one menu item rated very high (1), and all the distracters encountered prior to the highly rated item are rated low. The ANTICIPATE EA can have one of three outcomes (yes, no, maybe) which can increase, decrease or leave unchanged, the relevance of the item assessed. Figure 3 shows the sequence of assessments made by the model on these menus. Bold text has been used to highlight the high scent item.

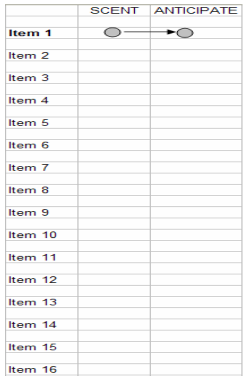

Figure 3: Assessments of items in a menu when the high scent item is in position 1 and the distracters are poor.

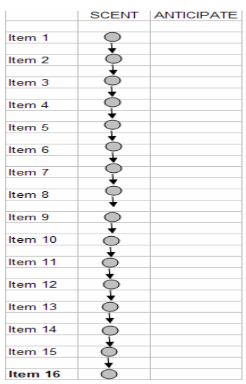

Figure 4: Assessments of items in a menu when the high scent item is in position 16 and all the distracters are poor.

A further two of the eight menus have the high scent item in the last position (16th) on the menu. On these menus the model chooses to do a SCENT assessment on each item and then, on reaching the bottom of the menu, stops immediately without conducting any ANTICIPATE assessments. This behaviour is illustrated in figure 4.

The second pattern in figure 1 shows an example of a participant assessing (nearly) all the items in the menu before refocusing on a previously assessed item and then selecting it. One might imagine that the agent considers the investment in the extra effort required to assess the rest of the list as worthwhile when considered against the probability that the high scent item is not the goal item. However, this is not an accurate way in which to conceptualise the behaviour of the model. Instead, this behaviour is most likely to occur when the previously encountered items have been discounted as irrelevant (scent = 5, R=0).

For the remaining four of these eight menus, the high scent goal item is in either position 7, 12, 14 or 15 in the menu while all the distracters have a low scent score. On these menus the model choose to do a SCENT assessment on every item in the menu before doing the ANTICIPATE assessment and stopping. This therefore provides us with an example of continued scanning before selection (figure 5).

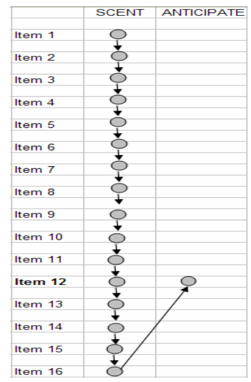

Figure 5: Assessments of items in the menu when the high scent item is in position 12.

Why the position of the goal item results in different behaviour. In order to understand how the position of the high scent item can result in these two different types of behaviour we will compare the behaviour of the model when the high scent item is in position two (as an example of occurring early in the menu) and in position twelve (as an example of occurring late in the menu) in more detail. In both examples, initially, all sixteen menu items are rated equally and all have a relevance (R) value of 0.06. In the first cycle, the EA that proposes assessing the scent of the first item in the menu is rated as having the highest efficiency due to it having the lowest cost. Consequently, the model assesses the first item which gets rated as a 5 (i.e. very low scent). As a result, the new R value of this item is set at 0. On the next cycle, the EA that proposes SCENT assessment on the second item in the list is the most efficient (due to the lower cost) so this item gets assessed. This behaviour continues until the model assesses the high scent item.

In menus where the high scent item occurs early on in the menu, the second item in the menu gets a R value of 0.5097 which raises the probability that this item will lead to the goal to 0.6220. On the following cycle (see figure 6), the R value of the high scent item (the black diamond) leads to an efficiency of 0.008 (the value of the low solid curve in figure 6 at 0.5097 on the x-axis) whilst the second best item (an item yet to be assessed – shown as a black square) has an R of 0.06 which results in an efficiency of 0.006 (the value of the dashed curve at 0.06 on the x-axis). Although the efficiencies of the two EAs are very similar, on is larger than the other rand this is what determines which EA is chosen. The grey diamonds in figures 6 & 7 show the relevance values of the other items in the menu. In figure 6, one item (that in position 1) has had its relevance reduced to zero and therefore lies at zero on the x-axis. All the other grey diamonds lie at 0.06 on the x-axis. Their height against the y-axis is immaterial.

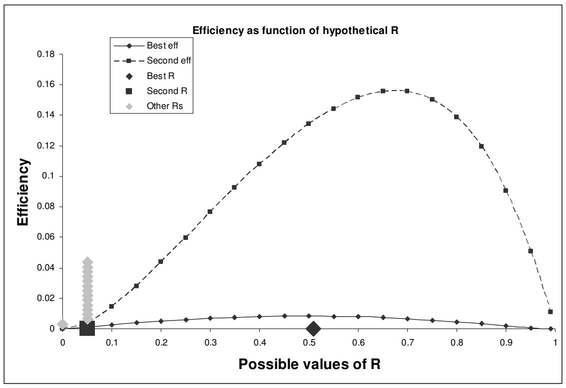

Figure 6: The R values for each item and the efficiency curves for best and second best EAs immediately after assessing the high scent item that occurs early in the menu.

In our example of a menu where the high scent item occurs late on in the menu, the relevance of each of the low scent items that have already been assessed falls to zero. When the model assesses the twelfth item, its relevance gets a value of 0.5097 which raises the probability that this item will lead to the goal to 0.6220. On the following cycle (see figure 7), the R value of the high scent item only has an efficiency of 0.005 (the value of the dashed curve in figure 7 at 0.5097 on the x-axis) whilst the item with the best efficiency (an item yet to be assessed) has an R of 0.05 which results in an efficiency of 0.006 (the value of the dashed curve at 0.06 on the x-axis). The result is that the model continues to assess each item in the menu until it reaches the bottom because the efficiency of conducting a SCENT assessment of a new item is greater than the efficiency of conducting the ANTICIPATE assessment on the high scent item in position twelve. This has the effect of slowly increasing the probability of the item in position twelve leading to the goal.

The reason this is different to the outcome shown in figure 6 is because the shapes of the two curves in figure 7 are slightly different due to the effect of the distracter items (or lack of them). The fact, in figure 6, that there are more distracter items still competing (i.e. with an R>0) has resulted in the peak of both curves shifting to the right which results in a flatter curve at the left-hand end and consequently lower efficiencies for items with low R values.

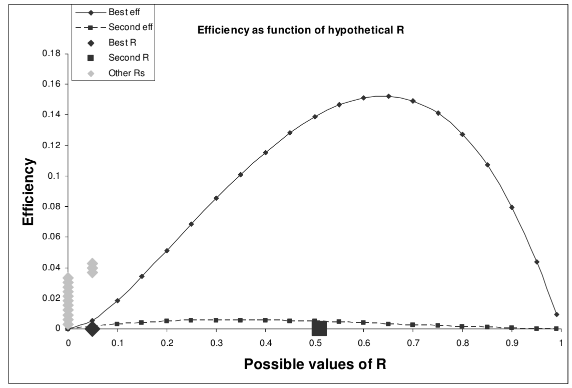

Figure 7: The R values for each item and the efficiency curves for best and second best EAs immediately after assessing the high scent item that occurs late in the menu.

Good goal, mediocre distracters. The remaining menus used in the empirical study (6/14) consisted of menus that were far more varied in terms of the pattern of scent. The goal item had a high scent rating (1) while the ratings of the distracters varied, both within and between menus, from 2 to 5 with a median of 3. On two of these six menus the high scent item is in position one on the menu. As in the previous examples where this is the case the model conducts a SCENT assessment of this item followed by an ANTICIPATE assessment and then stops. On a further menu, where the high scent item is in position three, the model conducts SCENT assessments of the first three items before conducting the ANTICIPATE assessment on the third item and then stops. Here again then we see examples of self-terminating behaviour. For these menus, the model does not ‘know’ that the rest of the menu has a more vaired scent pattern than those menus with the poor distracters (where all distracters were rated 5).

On the three remaining menus the high scent item is in position twelve (twice) or fourteen and the model produces patterns of repeated scanning similar to those of the participants. The following description of the model’s behaviour closely matches that of the participants’ scan pattern shown in figure 1. The model conducts a scent assessment on each of the items in the menu in turn (see figure 8). This is then followed by ANTICIPATE assessments on the highest scent items (those with a scent level of 2) on the upward scan. The model conducts another downward scan this time conducting ANTICIPATE assessments on those items assessed to have a scent level of 3. The final upward scan (shown by a dotted line) is inferred from the fact that, in order to complete the task, a model that included scent following would refocus its attention on the item in position 5 which had a scent level of 1 before selecting it.

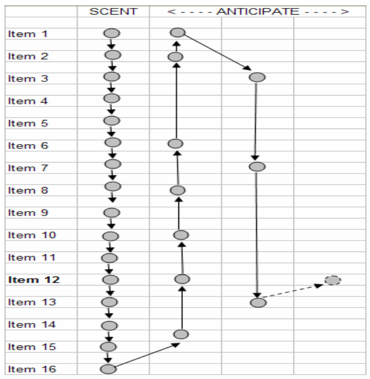

Figure 8: Assessments of the items in the mediocre menu when the high scent item is in position 12.

General Discussion

Comparisons of the traces of the model with the empirical data suggest that the model provides a good explanation of the cognitive processes involved in this task. This suggests that participants make an assessment of the relevance of a label to the current goal and then, together with the estimated relevance of previous items, choose to either i) select that item as the one that will lead to the goal, ii) conduct some further assessment of the current item, or iii) move on to another item and assess that. Which of these EAs is chosen is driven by the pattern of information scent that has been experienced so far.

The model provides us with an explanation of how and why the position of the goal and the quality of the distracter items affect the behaviour of the participants on the task. Regardless of the pattern of scent of the menu, our model predicts that the agent will tend to stop exploring the menu as soon as it comes across a menu item that has high information scent (self-terminates) if this is encountered early in the menu. On menus where there is one high scent item amongst a set of low scent items and the high scent item occurs later in the menu, the agent continues to assess the other items in the menu before conducting further assessment of the high scent item and finally selecting it. The model enables us to explain why we see these different patterns of behaviour on menus which have such similar patterns of information scent. This is due to the effect of the interdependence of the probability that each of the items will lead to the goal. The actual point on the menu at which the model swaps from one behaviour to the other is sensitive to a number of factors such as the length of the menu and the costs of the EAs. It would appear therefore that it is in the nature of exploratory behaviour that there are close calls which suggests that people can rationally do either behaviour and that a number of factors have an effect on the behaviour of participants exploring real menus.

When there are a number of distracter items that receive a similar level of scent assessment to that of the goal item (i.e. there are mediocre distracters) it is more difficult to identify the best item from amongst the distracter items and therefore additional cognitive effort is required in order to determine which item is most likely to lead to the goal. As a consequence, the agent is more likely to exhibit multi-scanning behaviour on these menus.

Although one might expect the data to show continued scanning of menus and multi-scanning one would not necessarily expect participants to self-terminate. The value of this research is that the model exhibits all three patterns of behaviour from a single set of parameters This suggests that no alternative strategies are necessary to explain variations in human behaviour on this task. The success of this model in providing an explanation of the eye-tracking data adds further support to the rational framework as a general model of exploration.

Acknowledgements

Thanks are due to Duncan Brumby and Andrew Howes for access to their menus and data and for comments on an earlier draft.

References

Brumby, D.P., & Howes, A. (2003). Adaptive decision making in menu search: The role of interdependence and past experience on link selection. *Proceedings Volume 2 of the* Conference *HCI 2003: Designing for society,* Bath, UK, 133-134.

Brumby, D.P. & Howes, A. (2004). Good Enough But I'll Just Check: Web-page Search As Attentional Refocusing. *Proceedings of the 6th International Conference on Cognitive Modelling*, Pittsburgh, USA.

Byrne, M. D., et al. (1999). Eye tracking the visual search of click-down menus. *Human Factors in Computing Systems: CHI 99*, New York: ACM Press.

Cox, A.L. (2002) *Exploratory Learning of Interactive Devices: What People Do and What People Learn.* Unpublished PhD, University of Hertfordshire, UK.

Pirolli, P. & Card, S., (1999) Information Foraging, *Psychological Review*, 104 (4), 645-675

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. & Cox, A.L., (2000) A New Rational Framework for Modelling Exploratory Device Learning … but does it fit with ACT-R? *Proceedings of the 7th ACT-R Workshop*, Carnegie Mellon University