Model of visual search and selection time in linear menus

Gilles Bailly, Antti Oulasvirta, Duncan P. Brumby, Andrew Howes

Model of Visual Search and Selection Time in Linear Menus

Gilles Bailly[1]   **Antti Oulasvirta**[1]   **Duncan P. Brumby**[2]   **Andrew Howes**[3]

[1]Max Planck Institute for Informatics and Saarland University
[2]UCL Interaction Centre, University College London
[3]School of Computer Science, University of Birmingham

ABSTRACT

This paper presents a novel mathematical model for visual search and selection time in linear menus. Assuming two visual search strategies, serial and directed, and a pointing sub-task, it captures the change of performance with five factors: 1) menu length, 2) menu organization, 3) target position, 4) absence/presence of target, and 5) practice. The novel aspect is that the model is expressed as probability density distribution of gaze, which allows for deriving total selection time. We present novel data that replicates and extends the Nielsen menu selection paradigm and uses eye-tracking and mouse tracking to confirm model predictions. The same parametrization yielded a high fit to both menu selection time and gaze distributions. The model has the potential to improve menu designs by helping designers identify more effective solutions without conducting empirical studies.

Author Keywords

Linear menus; User Performance; Mathematical predictive models; Visual search; Eye-tracking.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Menus are used on numerous applications and systems for presenting, organizing, and selecting commands. However, designers reportedly struggle with menu design [5]. While many research papers focus on novel interaction techniques (e.g. [38]), and others on human factors (e.g. [6, 9, 32, 33, 36]), surprisingly few develop predictive models of users' performance with menus [7, 8, 10, 12, 14, 16, 21, 28]. Predictive models are an efficient way of encapsulating scientific knowledge. They synthesize phenomena that are typically fragmented across several studies, and they capture subtle interactions among multiple factors in an empirically verifiable form. They can help designers to design more efficient menus by informing choices without expensive empirical studies. Once matured enough, models can be incorporated into interactive design tools [5, 16, 25].

Models of menu performance can be divided into two categories: cognitive simulations and mathematical models. Simulation models such as ACT-R models [3, 8] and EPIC models [19, 21, 27] explain the progression of search by reference to underlying cognitive processes, such as perception, attention, and memory. Mathematical models such as the SDP model [12, 14] are nonlinear regression equations that predict selection time as a function of external factors such as menu length, target position and practice. These models expose fewer details about the process but may be less complex and more straightforward to apply.

Our goal is to advance the scope and validity of mathematical models of *linear menus*, a widely used menu type in desktop applications and consumer electronics. We assume that tasks, including both searching and pointing, are performed using one of a number of strategies. In practice, the strategy space for perceptual/motor tasks, even very simple ones, is known to be large [23, 39] but here we explore a few key points in the space. We focus on two searching strategies: 1) Directed search makes use of knowledge, including knowledge of locations acquired through practice, alphabetic and semantic knowledge, to programme saccadic eye movements that jump to the vicinity of the target location or directly to its location. 2) Serial search starts with the first menu item and moves from one item to the next until the target is located or the menu is exhausted. In practice a user would combine elements of these strategies depending on the reliability of available of knowledge. For example, directed search might be combined with serial search in proportion to the reliability of location memories and the available relevance scores. In some situations location knowledge will be highly reliable and in others it will be the relevance scores that are more reliable. Finally, 3) we also explore two pointing strategies: *Find-and-point* and *Track-and-point*.

In our current model these strategies are expressed as probability distributions of gaze. A distinct advantage of this model formulation is that it allows deriving both distributions for gaze points on the menu as well as selection time from the same parametrization. These predictions can be checked against eye tracking data.

To parametrize and evaluate the model, we collected an extensive dataset. Although the constituent processes of menu selection have been heavily studied in psychology, the menu selection task has been hard to study for the purposes of modeling. Multiple factors must be controlled, the apparatus like eye-trackers are complex, and the studies must be long enough to gauge practice effects. Our model is informed by

a study that replicates and extends the Nielsen [36] menu selection task in order to cover more factors. The experimental design controls 1) menu length, 2) menu organization, 3) target position, 4) absence/presence of target, and 5) practice. These factors were chosen to extend the coverage of modeling to a larger number of empirically reliable findings reported in the literature. Particularly important is the effect of menu organization, which we cover with three frequent layouts: alphabetical, semantic, and unordered. Moreover, we reveal a *last item effect*: items that are positioned toward the end of the menu are faster to select. The data confirms some existing findings and yields new observations of subtle interactions among the factors that our model tries to capture.

GOALS AND RELATED WORK

In our view, an ideal predictive model covers phenomena which are replicated, shown to bear large effect sizes, concern features that designers can affect, and are widespread in real designs. We visited empirical findings on linear menu search and gathered a list of six phenomena which serves as the goal for our modeling efforts. We review them and discuss what is left out of our model due to this scoping. We also briefly review how existing models address these phenomena.

Menu Performance Phenomena

1. Menu length: Users are faster at searching menus with fewer items [2, 9, 36]. More precisely, search time to the same target position increases as the number of items in the menu increase. This effect can be explained by the fact that, on average, it will require more time to localize a random target in a longer menu [1, 14, 21, 38].

2. Menu organization: The organization of items in the menu affects selection time. In particular, semantic and alphabetic organizations of items are faster than unordered layouts [10, 18, 32, 33]. This is explained by learning in visual search: experienced users learn to skip items in the menu. The relationship between alphabetic and semantic organization is more nuanced. For targets with unknown labels, semantic organization is faster than alphabetic organization [32]. However, for target with known labels the evidence is mixed: One study found no difference [32], but two others suggest that alphabetic is faster [10, 33].

3. Target position: Users are faster at selecting targets that are closer to the top of the menu [9, 14]. This effect is large and only partially explained by pointing time. Visual search strategies can explain the effect [37]: Users can follow a *serial search* strategy, reading items one by one from the top of the list until reaching the target [9, 21, 29]. Users can use a mix of serial and random search (with or without replacement) [20, 21]. In both cases, search time is a linear function of target position. Moreover, the second item can sometimes be faster to select [8, 9]. Finally, we will reveal below that items that are positioned toward the end of the menu are faster to select (Fig. 2). We call this phenomenon *last item effect*).

4. Absent target: Scanning a menu that does not contain the target takes longer than scanning one with it. This can be explained by exhaustive visual search: Novice users have to scan all items to ensure the target is absent.

5. Practice: Performance improves with more experience with the same menu [14, 26, 40]. One explanation is that when the layout is stable, users remember the location of items and can directly glance at them [40]. Moreover, three studies [10, 32, 33] show that the effect of organization disappears with practice [8].

6. Mouse control: Users mainly exhibit two strategies to coordinate its movement with the eyes: 1) a single move once the target item has been located by visual search and 2) the tracking strategy where the cursor trails the eye gaze [9]. The latter one decreases time and the contribution of pointing, because the cursor is already closer to the target. The mouse cursor can also be used to tag an item while the eyes are free to move elsewhere and access additional knowledge [15].

Our decision to focus on these effects leaves out some other phenomena for future work. One such effect is naming: Item labels that are clearly associated with the target goal are quicker to localize [6, 33]. Another one is distractors: Searching for a number among letter distractors is faster than searching for a letter [2]. The density of items impacts also visual search [17, 35]. Related to this is the saliency effect: Items that are salient in terms of color or size can be easier to localize. We believe that our scope is justified on the basis that all six effects have been replicated and bear large effects.

Overview of Previous Models

The main difference between mathematical and simulation models is that the latter have a richer description of visual search strategies and assume it to dominate total selection time. Our goal is to incorporate their findings to a mathematical model.

Mathematical models: The models pre-dating the SDP model assume a linear relationship between item position and search time [28]. The *SDP* model (Search-Decision-Pointing) [12, 14] is a regression model using four variables and seven parameters to predict total selection time. Contrary to the simulation models, it has only a random model of search and it places more emphasis on decision and pointing time. The four predictors are: *Number of items*, *item frequency*, *spatial consistency* and *item position*.

SDP covers some of the phenomena mentioned above: 1) the **S**earch component predicts that the time to localize an item increases linearly with the number of items in the menu. However, the position effect is only explained by pointing time, not by localization time. 2) The **D**ecision component predicts that the time to decide among items depends on the "entropy" of their relative frequencies in previous selections. It is given by the Hick-Hyman law. 3) The **P**ointing component is based on Fitts' law [30] and predicts that items closer to the top are faster to select. Furthermore, mouse movement starts *after* search. 4) Finally, an *expertise* scalar modulates the components by the number of repetitions with an item. With practice, performance shifts from being dominated by search (linear) to decision (logarithmic). 6) The model includes a scalar "predictability" variable (1: unchanging, 0: constantly changing order) that predicts a detrimental effect of spatial consistency. With these assumptions, the

SDP model has been extended to scrolling lists and hierarchies [12], square and circular layouts [1] and interaction with constrained devices [13].

Simulation models mainly refer to visual search as the explanatory factor for the above effects. *EPIC* (Executive Process Interactive Control) [27] consists of a production-rule cognitive processor and perceptual motor peripherals. Their execution is limited by temporal and capacity limitations posed by the cognitive infrastructure [19, 21, 22]. Four strategies are distinguished for menus [21]: serial search (one menu item processed at a time), parallel search (many items processed at the same time), random search, and systematic search. The last two are combination of the others. Data suggested that parallel search with both random and systematic search matched well observed data.

ACT-R/PM [8] extends ACT-R [3], which is also a production rule architecture. It differs from EPIC in the details of the visual search process [35]. ACT-R/PM posits a systematic, top-to-bottom search with eye fixations on menu items that share features with the target item [3]. In constrast, EPIC posits a 50/50 blend of random and systematic search with eye fixations determined by the number of menu items visible in the fovea. However, a study found that neither EPIC nor ACT-R is a good characterization users' behavior [9]. The assumptions concerning mouse movements also differ. EPIC predicts, similar to SDP, that there is a single aimed mouse movement from the initial position to the target item once the target has been located. Timing of this movement should be governed by Fitts' law [30]. ACT-R, on the other hand, predicts that the mouse should "trail" the eyes such that once the target is located, there is an approximately constant and short distance to the target. This predicts multiple mouse movements that are correlated with the number of eye movements.

MODEL OVERVIEW

The model has four input variables: target position (1-16), menu length (8-16), menu organization (unordered, alphabetical, semantic), and number of previous encounters with the item (1-12). The model is in a mathematical form that allows deriving both total menu selection time and gaze distribution. The assumption is that the total time to select an item consists of time spent in two consecutive subtasks–search and pointing–and both are affected by learning (Figure 1):

- **Search** uses the recruitment of eye gaze and memory recall to identify the position of the target in the menu by scanning it. We follow previous simulation models [20, 21] and assume two search strategies:

    - **Serial search** consists of top-to-bottom reading of items that, with practice, allows more skipping. This is modeled as a uniform distribution from the beginning of the menu to the target.

    - **Directed search** consists of a direct attempt at moving the eyes on top of the target. At first such attempts are random, as the user tries to guess the location, but with more exposure they become more accurate. We model directed search as a normalized Gaussian distribution centered on the target.
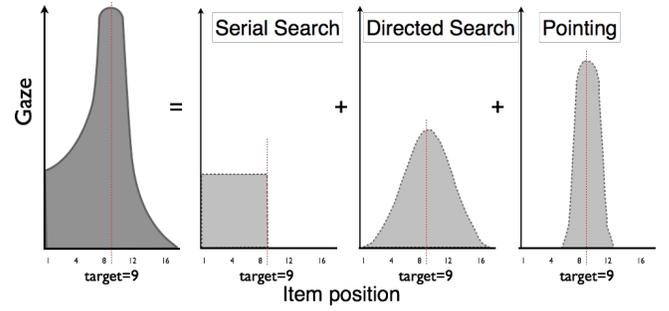

Figure 1. We predict total selection time as a gaze distribution that is the sum of three components: Serial search, Directed search and Pointing.

- **Pointing**: After localizing the target, the target is selected by pointing. Pointing time is modeled with Fitts' law, like in previous models. However, like [8], we assume two starting locations of the cursor: 1) the first at the top of the menu, 2) the other from a constant distance in the vicinity of the target (trailing).

- **Learning**: Our learning model is a power law of practice model [34] which is parametrized for each search component. The negative exponential form of the law implies that largest gains are achieved in the first repetitions, and the gains diminish rapidly after that.

Total Selection Time

Total selection time T is the sum of time spent in serial search T_s, directed search T_d, and pointing T_p:

$$T(\theta) = T_s(\theta) + T_d(\theta) + T_p(\theta) \qquad (1)$$

where $\theta = (l, t, P)$ is a vector containing the four input variables: length of the menu l, position of the target t, and a vector giving the number of previous encounters with the items in the menu P (practice).

An original aspect is to explain *where* (e.g., on which item) the eyes spend time. The total time for selecting target t is the sum of the gaze time required for each item i of the menu. The relationship between total time and the gaze distribution for each component is:

$$T(\theta) = \frac{1}{R} \sum_{i=0}^{i=l} G_s(i, \theta) + G_d(i, \theta) + G_p(i, \theta) \qquad (2)$$

where G_s, G_d, G_p estimate the numbers of gaze points (serial, directed, and pointing) captured using an eye-tracker with a sampling rate of R. In our studies, $R = 50Hz$.

Serial search

Serial search refers to the top-to-bottom inspection of items. Items after the target are not visited. With practice, users not only skip more items but also spend less time on each visited item. For this reason, the number of gaze points on each item in a menu is not constant but decreases as a function of repetitions. The number of gaze points for i when looking for target t is:

$$G_s(i, \theta) = \begin{cases} a_s \times \exp(-b_s \times p_i) + c_s & \text{if } i \leq t \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where a_s, b_s, c_s are the parameters of the learning model, while p_i is the number of previous encounters with item i.

Directed search

In directed search the user tries to glance directly at the target. Attempts are initially more random, but the spread of gaze distribution decreases with more repetitions. As experienced users have memory about the general location of the target, they can glance at the neighborhood of the target and, eventually, directly move the eyes to it [11]. As in serial search, we assume that experienced users also skip more items and spend less time reading them than novices.

We thus model this distribution of gaze with a normal distribution centered on the target item t. Probability for item i in the menu is given by its probability density function $\psi_{t,\sigma_d^2}(i)$ modulated by the power law of learning:

$$G_d(i,\theta) = \psi_{t,\sigma_d^2}(i) \times (a_d \times \exp(-b_d \times p_i) + c_d) \quad (4)$$

where σ_d is the standard deviation reflecting the spread of the search area. a_d, b_d and b_d are the empirical parameters of the learning model. Note that the sum in Eq. 2 ranges from the first item to the end of the menu ($i \in [1, l]$), but not longer, because we assume that users only search inside the menu.

The spread σ_d reflects the accuracy of memory-directed search and depends on the number of previous encounters p_t with the target. The amplitude of the distribution reflects the amount of time spent in the search area and is affected by the power law of practice. We assume that the spread σ_d is also affected by menu length: Longer menus have more uncertainty and users are more likely to visit more items:

$$\sigma_d(i) = l \times a_\sigma \exp(-b_\sigma \times p_i) + c_\sigma \quad (5)$$

where a_σ, b_σ and c_σ are the parameters of the learning model. Note that two learning components affect the gaze distribution: The first affects the tendency to engage in directed search (amplitude), the second the accuracy of perceptual memory (spread).

Pointing

Pointing is the time the eyes spend on the target to make the selection after it has been localized. The pointing time is predicted by Fitts' law [30]:

$$T_p = a + b \times log_2(1 + ID) \;\; with \;\; ID = \frac{D}{W} \quad (6)$$

where ID is the index of difficulty. In our model, D_t is the distance from the cursor position to the target position when starting the pointing task. Evidence [9] suggests that the cursor sometimes trails the eye gaze and can be in the vicinity of the target when this one is localized. However, the distribution of these two strategies (starting from the top vs. starting in the vicinity) is not known. The average distance D_t is thus defined with a weighing parameter α:

$$D_t = \alpha \times d_t + (1 - \alpha) \times d_0 \quad (7)$$

where α reflects the relative usage of the two strategies; $d_t = t * W$ is the distance from the top of the menu to the target

and d_0 the distance from the cursor to the gaze when using the trail strategy. We assume that $d_0 = 0$.

$$T_p = a + b \times \log_2(1 + \alpha \times t) \quad (8)$$

We can thus estimate G_p. However, to account for the fact that users can see 2-5 items in the neighborhood of the gaze point, we revisit the equation 8. We add a normal distribution $\psi_{t,1.1}(i)$ that covers the dispersion of gaze in the items around the target:

$$G_p(i,\theta) = a_p + b_p \times \log(1 + \alpha \times t) \times \psi_{t,1.1}(i) \quad (9)$$

where $\psi_{t,1.1}$ is a normal distribution centered around the target t with a standard deviation equal to 1.1. This reflects that 95% of the gazes are in the users' (extended) fovea.

Absent items

We also provide a simple model for deciding that an item is not in the menu. We assume that search time here only depends on menu length. If users do not know the menu content ($p = 0$), they inspect *all* items. They then become quicker at detecting missing items. The model is:

$$T_a(\theta) = a_a \times exp(-b_a \times P_a) + c_m \quad (10)$$

where a_a, b_a, c_a are three components of the learning model. We expect here having a large b_a reflects that users quickly acquire knowledge about absent targets.

DATA COLLECTION

The experimental design is based on and extends Nielsen [36]. In the experiment, a label is shown and the participant must click the corresponding item as quickly as possible. Distinctive properties of the method include:

- eye fixation and mouse movement data are collected to understand visual search and pointing behavior (cf. [9])
- item selections are repeated multiple times to understand practice effects
- three different organizations of the items are included (unordered, alphabetical, semantic).

Participants, Apparatus and Setup

Twenty one participants (13 females), ages 18 to 26 years (mean=21.1, σ=3.54), were recruited by email list advertisements, and paid 45 euros for participation. All participants had normal or corrected-to-normal vision, were familiar with WIMP interfaces, and used menus regularly.

The experiment was conducted on a windows PC and a 20 LCD display at 1280x960 resolution. A traditional optical mouse was used. All Participants used the same sensitivity and acceleration of the mouse.

We used an 1750 Tobii eye-tracker. The sampling data rate is 50Hz (raw eye movement data points every 20 ms); the latency was 20ms and the spatial resolution is 0.25 degrees (0.19cm on the screen) while the height of an item is 0.75 cm. The distance of the users' eyes from the screen was 65cm. It means that participants can simultaneously focus on 3 items if we consider a fovea of 2 degrees.

Experimental Design

The experiment used within-subjects design with four independent variables: *Menu length*, *Menu organization*, *Practice*, and *Target location* as the main factors. Menus contain 8, 12 or 16 items that are common menu lengths in current applications [31]. We tested three menu organizations: Unordered, Alphabetical and Semantic organization. In the semantic organization, logical groups contained 4 items for all menu lengths. 4 items were chosen because it is the average number of items per logical groups according to [4]. Logical groups were separated by an horizontal line (separator) like traditional linear menus. To investigate the impact of practice, participants carried out 12 blocks per menu. Each item is selected the same number of times. Additionally, we added a missing item in each block. The order of menu length as well as menu organization was counter-balanced between participants. Participants searched one menu organization per session. Menu length was counter-balanced between menu organization. For each layout, participants performed at least 4 blocks. Finally, the order of items was randomized within block. In summary, we have: 22 participants x 3 Menu configurations x 3 Menu Length x 12 blocks (depending on menu length) x (8-12-16 items) = 39,564 selections.

Task and materials

The task consists of selecting the target as fast as possible. Each trial starts with a single button on the screen. As soon as the cursor enters in the button, the stimulus (label of the target item) appears just above the button (to avoid possible occlusion). The selection task started as soon as participants pressed the button. The label then disappears and the menu appears. The task finishes when participants activate the *correct* target and then the menu disappears. In the case of missing items, it finishes when participants press the space bar.

We identified several kinds of stimuli in the literature on menus: the explicit target name, a synonym, the definition of the target or an icon [6, 33]. In this experiment, we used the explicit target name to avoid confound effect.

The same set of items was used for each menu organization. The set of items was changed for each menu length. The set of words that we used is a subset of the list that can be found in [18] with minor modifications: no items in the menu have more than 3 letters length difference. This modification was used to reduce potential saliency effects.

To assess the validity of semantic groupings, seven participants were asked to rate the semantic relatedness (0 to 100) of 120 pairs. Item-pair order was scrambled and the respondents did not know if they originated from a semantic menu or not. We found that items that are in the same semantic group had average relatedness of 44.4 (σ = 26.0) whereas other items had 16.0 (σ=10.0). A two-sample unequal variance t-test found this difference to be statistically significant, $t(47) = -6.828, p < 0.0001$.

Procedure

After introduction to the task, the eye tracker was calibrated. During the experiment, participants took a break every 50 selections. We forced the minimum duration of the break to 1 minute to avoid fatigue. Breaks between conditions were used to re-calibrate the eye-tracker. The experiment was split into 3 sessions of 90 minutes each. The time between two sessions was from 0 to 7 days depending of the schedule of the participants. Participants were instructed to select items as fast as possible. They were not instructed about the layout and the length of the menu they will use. Finally, before the experiment, participants performed a calibration task consisting of selecting a target in a menu with a single item.

RESULTS

To preprocess the data, we removed trials in which selection times were greater than 3 SDs from the mean and trials in which the participant did not correctly select the target on the first attempt. Using these criteria, 3.83% (1,515 out of 39,564) of trials were excluded. In all subsequent analysis, one block has 3 repetitions.

Inferential statistics

To gain an overview of the data we carried out a series independent Organization × Practice × Target position ANOVAs on the selection time data. Table 1 provides a summary of the results of this analysis. To simplify our discussion, we focus on the effects that the experiment was designed to replicate and extend: (1) practice effect, (2) target position effect, (3) absent target effect, (4) menu length effect, and (5) menu organization effect. These effects are shown in Figure 2.

Practice effect. Table 1 shows that there was a consistent effect of Block, reflecting that users became faster at locating the target with repeated practice at searching the same menu. Table 1 also shows that there were significant interactions between Block and the other factors.

Target position effect. There was a robust effect of Target position on search time (see Table 1). To understand this effect, Figure 2 plots selection time as a function of target position for different menu lengths for the first (2-a) and last block (2-b) of trials. It can be seen that during the first block, selection time tends to increase with target position. By the last block of trials, this target position effect has decreased (i.e., the curve has flattened).

Absent target effect. Recall that on some trials the menu did not contain the target (an absent target trial). On these trials the participant had to press the space bar to indicate that it was absent. Response times are shown in the Figure 2a -b (at 0 on the x-axis). It can be seen that participants were initially very slow at making these absent target decisions during the first block of trials (particularly for 12- and 16-item menus). By the last block of trials, participants were very quick at indicating when there was an absent target.

Length	Organization (O)	Practice (P)	Target (T)	OxP	OxT	PxT	OxPxT
8	**	***	***	n.s.	***	*	**
12	***	***	***	***	***	***	n.s.
16	***	***	***	***	***	***	*

Table 1. Summary of main effects and interactions for search time data.
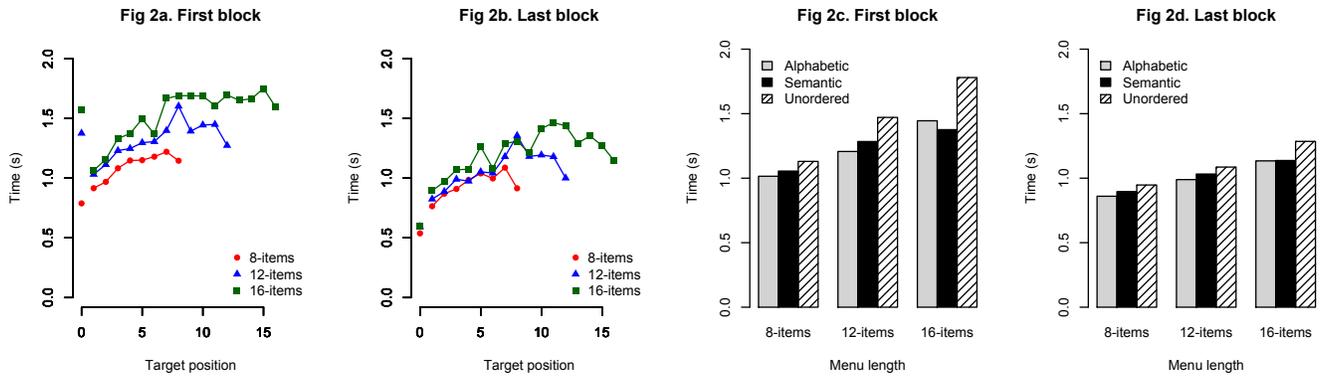**Note: * p < .05, ** p < .01, *** p < .001**

Figure 2. The effects of target position (left panels) and organization, and practice interaction (right panels) on selection time.

Menu length effect. The data replicate Nilsen's menu length effect: selection times were longer for targets in a given position when the menu contained more items. To test this effect, we performed a Length × Organization × Practice × Target-position ANOVA on a subset of the data: trials in which the target was in position 1 to 8. As expected, there was a significant main effect of menu length, $F_{2,40} = 70.44, p < .001$. However, there was also a significant Length × Block interaction, $F_{6,120} = 8.84, p < .001$. With practice, the effect decreases from the first to the last block (Figure 2a b).

Menu organization effect. Figure 2c d shows the menu organization effect. Participants were faster at selecting targets when the menu had either an alphabetic or semantic organization compared to when it had an unordered organization. Reflecting this observation, statistical analysis shows that there was a consistent main effect of menu organization across (see Table 1). There was also significant organization × block interactions for the longer 12-item and 16-item menus. As can be seen in Figure 2, these interactions reflect that with practice the menu organization effect diminished.

In contrast with the Nilsen experiment, we observed a *last item effect*: Items that are positioned toward the end of the menu are faster to select. It is the opposite of what would be expected by the Fitts' law. We return to this effect later in the paper and show that our model explains this effect as well as the previous effects and their interaction with practice.

Observations of Pointing Strategies

We here report observations for pointing. We return to gaze distributions in the Modeling section.

The model of pointing in Eq. 9 assumes that there are two strategies of mouse control and the proportion of these is regulated by the parameter α. Figure 3 shows two examples of cursor controls for the same target position: on the left trailing behavior, on the right the single move behavior. To better understand cursor control and to derive α, we looked at the position of the cursor upon localizing the target. To identify this moment, we used a simple heuristic: we assumed localization to take place at the beginning of the last fixation to the target item (Figure 3). This estimate is conservative, because it does not take into account the time needed to identify
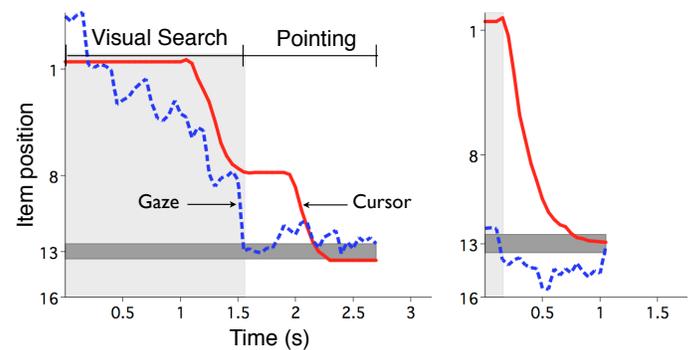

Figure 3. Two cursor control strategies for selecting the target 13 (dark grey). Left: The cursor (solid) trails the eyes gaze (dash). Right: A single mouse move when the target has been localized.
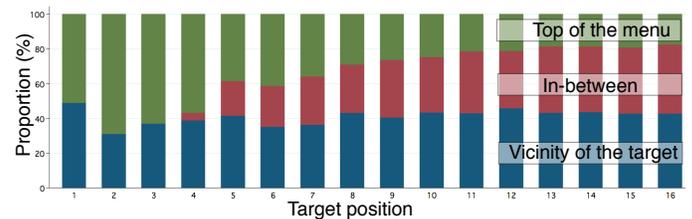

Figure 4. Proportion of three cursor movement strategies: trailing (on the target +/- 0.5 items), single move (top of the menu; +/- 0.5 items), and intermediate (between).

the target. Figure 4 shows the relative proportion of mouse strategy according to the target position. Data confirm our assumption that pointing and search are not always sequential.

MODELING

In this section, we estimate parameters and report model fit. Unlike previous models of menu selection, our approach allows assessment at three levels of granularity: *Selection time for menu*, *Selection time for targets*, and *Gaze distribution*. We used trials containing no missing gaze data (85% of the trials). This is important because we derive selection time from the gaze distribution (integration). Two participants were removed due to incompatibility with the eyes tracker (less than 65% of valid trials). It results in 19690 trials aggregated between participants.

Parameter Estimation

The model consists of 4 independent variables. For each menu organization, we have 12 parameters, 5 of which are shared by all menu organizations. Five parameters were estimated by analysis of user performance in selected conditions. The rest were solved using ordinary least squares (OLS), but insisting on theoretically sound value ranges. Note that we assume that both visual search strategies depend on menu organization: Unordered (U), Alphabetic (A), Semantic (S) [10, 18, 32, 33]. The parameters of Eq. 3, 4, 5 are thus fitted separately for each menu type.

Serial seach: Parameters a_s and b_s were obtained by OLS. For the three menu types, this yielded a_s =[U: 2.04; A: 1.48; S: 1.05] and b_s =[U: 0.15; A: 0.07; S: 0.09]. Second, we assume c_s is the same for every menu. This reflects the observation that an expert user can directly glance at the target *regardless* of the menu organization. We estimated c_s based on our Fitts' law calibration task to be $c_s = 0.2$. Notice that all parameters (a_s, b_s, c_s) are positive, describing that users abandon serial search with practice. We also note that they spend more time in serial search with the unordered menu ($a = 2.1$) that with the two other organizations [1.48, 1.05].

Directed search: The two parameters a_d =[U: 4.83; A: 4.28; S: 4.36], b_d =[U: 0.03; A: 0.03; S: 0.03] of the power law of practice, as well as the three describing the spread of directed search, σ_a =[U: 0.14; A: 0.14; S: 0.27], σ_b =[U: 4e-9; A: 1e-9; S: 5e-9] and σ_c =[U: 1.2; A: 1.5; S: 1.24], were obtained using OLS. That all parameters are positive reflects that not only users spend less time on each item but they also reduce the search area. Additionally, b_d, σ_b and σ_b are very similar whatever the menu organization indicating that the organization only impacts the initial amplitude and the initial size of search area. Finally, $c_d = c_s = 0.2$.

Pointing: We used data from the Fitts' law calibration task to estimate the parameters of Eq. 9 as $a_p = 10.3$ and $b_p = 4.8$. We consulted Figure 4 to set $\alpha = 0.4$. These parameters are organization-independent.

Absent target: We used OLS on the data of absent items and set the parameters to: $a_m = 3.75$, $b_m = 1.14$, and $c_m = 23.8$. We assume these parameters to be independent of menu organization, because every item is read.

Pre-search gaze

Figure 5 shows deployment of gaze on the stimulus before searching the menu. This occurs because users in the study processed the stimulus even a couple of milliseconds after having opened the menu. To cover this in the model, we added a component *Pre-search*: a normal distribution centered around the stimulus location (above the menu). It reflects the dispersion of gaze around the stimulus. We assume this component to be constant:

$$G_{pre}(i) = a_{pre} \times \psi_{S,\sigma_{pre}}(i) \qquad (11)$$

where $a_{pre} = 6$, $\sigma_p re = 1.0$ were defined using OLS. The two parameters are, respectively, the amplitude of the normal distribution and the standard deviation (spread). $S = -0.5$ is the position of the stimulus.
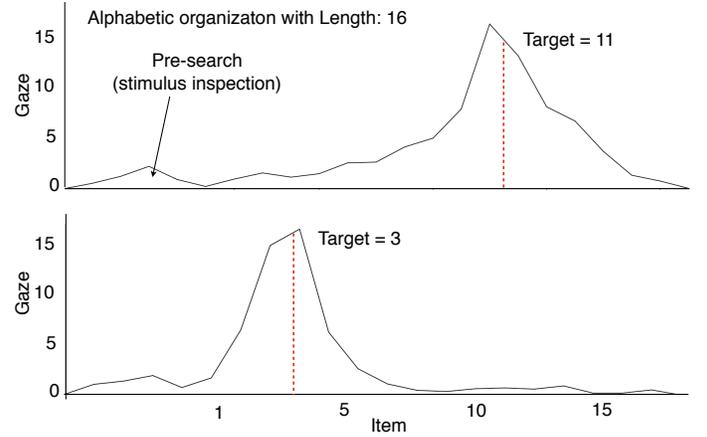

Figure 5. Gaze distribution for the target 3 and 11 aggregated across users. The plot shows that gaze occurs also in the vicinity of the stimulus (above the menu).

Model Fit

We report fit for three levels of analysis: *Selection time for Menu*, *Selection time for Target* and *Gaze Distribution*. The parameter values reported above were utilized for all three levels. Models fit are based on the whole data set. We also report at the end of this section a 2-fold cross-validation to avoid possible overfitting. All predictions for the whole design of the experiment are given in Supplementary Materials.

Selection time averaged for menu: Empirical data matched the predictions well for organization, length, and practice, as shown in Figure 6. Model fit was $R^2 = 0.95$ ($pred = 0.93 \times emp - 0.01$) for Unordered, $R^2 = 0.89$ for Alphabetic ($pred = 1.09 \times emp - 0.09$) and $R^2 = 0.78$ for Semantic ($pred = 1.84 \times emp - 0.74$).

Selection time per target: Model fit was $R^2 = 0.76$ for Alphabetic ($pred = 0.76 \times emp + 0.24$), 0.71 for Unordered ($pred = 0.87 \times emp + 0.25$) and $R^2 = 0.61$ for Semantic ($pred = 0.43 \times emp + 0.54$). While the absolute value is inferior to the menu-level prediction, many more observations must be predicted [1]. Figure 7 shows a sample of observed vs. predicted values for the different conditions. The model predicts the known effects of target position, menu length, and practice. It also predicts that selection time does not linearly increase with target position. More surprisingly, it predicts the decrease in selection time for the last item (last item effect). These two last effects can be explained by the Eq. 4, which models gaze distribution by a normal distribution centered around the target. When the target is the last in the menu, the predicted gaze distribution is effectively "cut" for items after the menu, lowering the total amount of gaze for that target item. This reflects the observation that users hardly look at the area below the menu. We also observe a limitation of our model for the semantic organization. Selection time increases quickly as a function of target position, especially for the first block. This can be explained by the serial search

---

[1] We reanalyzed Cockburn et al.'s data [14] for selection time per target. Their fit is at $R^2 = 0.42$ for this task. The analysis was based on data available at **www.cosc.canterbury.ac.nz/~andy/publs/chi07/**.

component. We assumed that users inspect all items from the top of the menu to the target. We hypothesis that users can skip more items with the semantic menus by using logical groups and separators.

Gaze Distribution. This prediction task has 13,800 points to predict (gaze for each item according to target, length, organization and practice). However, empirical data match the model surprisingly well, yielding a fit of $R^2 = 0.90$ for both Unordered ($pred = 0.90 \times emp + 0.05$) and Alphabetic ($pred = 0.85 \times emp + 0.40$) and $R^2 = 0.76$ for Semantic ($pred = 0.81 \times emp + 0.92$). One sample is shown Figure 8. We provide more examples in Supplementary Materials.
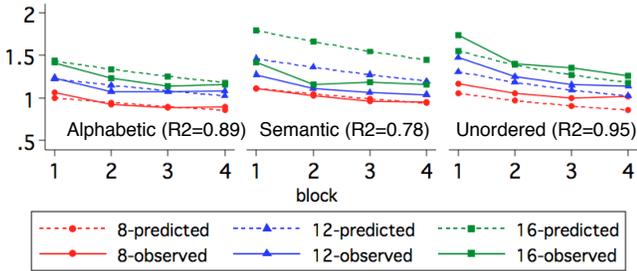

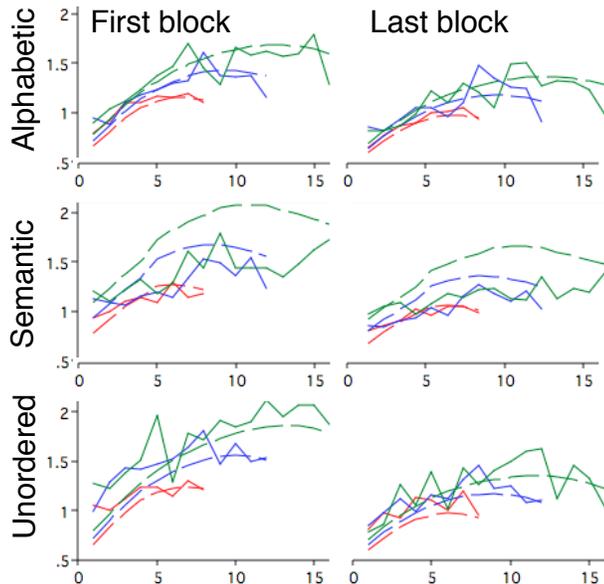**Figure 6. Observed vs. predicted menu selection times by menu organization, menu length an practice.**

Figure 7. Observed (solid) vs. predicted (dash) menu selection time per target, organization length for the first and last block
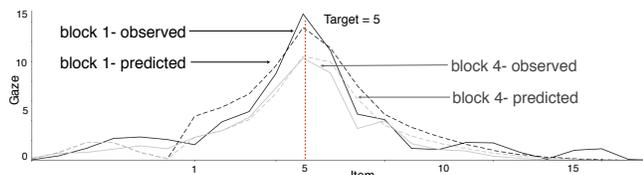

Figure 8. Actual (solid) and predicted (dash) gaze distribution aggregated accross all participants for the unordered organization at the block 1 (black) and block 4 (grey).

Cross-validation

We performed a 2-fold cross-validation analysis to test the generalizability of our model. We calibrated the model parameters on half of the participants and tested it on the others. The R-squared is still high at the menu level (A: 0.91; S: 0.86; U: 0.87), target level (A: 0.64; S: 0.52; U: 0.56) and item level (A:0.85; S: 0.81; U:79).

DISCUSSION

The present paper advances the understanding of menu performance by providing a mathematical model that better accounts for pointing behavior and visual search. It achieves this by combining assumptions about search and pointing strategies that have been identified by previous literature but never modeled in a single model.

We proposed a novel formalization to capture these behaviors. Instead of modeling sequences of elementary operations, like in cognitive simulations, or a linear sum of time spent in sub-tasks, we model the distribution of gaze. The benefit is that the area under the curve is an estimate of the total selection time. Thus, the same model can be used for predicting both dependent variables. The data support the assertion that modeling visual search distributions is a promising approach to mathematical modeling of linear menu performance. The predicted distributions not only yield a good fit but they provide insight into the process of search.

We conclude with a critical appraisal of the model in the light of six criteria for predictive models by Jacobs and Grainger [24]:

1. **Plausibility**: Are the assumptions of the model psychologically plausible?
2. **Explanation**: Is the theoretical explanation reasonable and consistent with what is known?
3. **Interpretability**: Do the model and its parameters make sense?
4. **Descriptive adequacy**: Does the model provide a good description of the observed data?
5. **Complexity**: Does the model capture the phenomenon in the simplest manner?
6. **Generalizability**: Does the model predict the characteristics of data that will be observed in the future?

Plausibility. We modeled pointing time with Fitts' law as previously, but with the additional assumption that the distance to target depends on the selected pointing strategy. One strategy starts from the top, the other trails the eyes. The two strategies were included in previous work but never in a single model [9]. Visual search was also modeled by assuming two strategies: serial and directed search. The direct search component makes the assumption that search is random in the beginning of experience, but becomes gradually more accurate. It thus subsumes a model of random search assumed in previous works (e.g [14]). Unlike the SDP model [14], our model does not assume a separate component for decision. More work is needed to understand the role of decision processes in menu selection.

Explanation. The model increases our ability to cover important design factors by a single model. The model covers the effects of menu organization, absent item, and target position, as well as the previously modeled factors of menu length and experience. The model also describes the "last item effect" that was not reported in the literature and is counter-intuitive.

Interpretability. Our modeling assumptions come from previous work. Moreover, the model increases the interpretability of mathematical models in this area by linking selection time to gaze. The model makes predictions of gaze distribution that are verifiable with eye-tracking data.

Descriptive adequacy. The model not only provides a high fit at the menu level but also at the target and gaze level. This multi-level approach decreases the risk of over-fitting and better describes the collected data. It is the main difference with previous models.

Complexity. Our model is more complex if measured by the number of free parameters, but the ratio of parameters against explained observation points favors our model. A single parametrization was used to describe 13800 gaze points, 432 target-level selections, and 36 menu-level selections.

Generalizability. To test the generalizability of our model, we provided a cross-validation result that were mostly favorable to our model. However, novel experiments should be carried out to test generalizability to completely different contexts, for example including other menu organizations, layouts, etc.

We see two opportunities to develop the model further. First, more work is needed to cover other common features of linear menu designs (e.g., icons) as well as more complex menu *systems* that include hotkeys, toolbars, and hierarchical menus. This means that semantic relatedness among items must be modeled to predict their effect on navigation and exploration behavior [16]. Second, although we are now better able to model search strategies, our model only predicts averaged behavior. This unfortunately can make the approach "fragile" to the sample of users. Were the distribution of strategies to change in the sample, the averaged prediction may not hold.

Future work also needs to consider a more thorough exploration of the strategy space implied by the statistical distribution of menu relevance and the cognitive constraints. For example, Smith et al. [39] reported a multi-dimensional analysis of the space of possible typing strategies and showed that the assumption that people adopt rational strategies can be used to predict performance. We are currently building a similar analysis of menu search that is aimed at revealing the details of how directed and serial search strategies are combined depending on (1) the availability and precision of relevance and location knowledge and (2) the ecological distribution of relevance and organizational factors in real-world menus.

REFERENCES

1. Ahlström, D., Cockburn, A., Gutwin, C., and Irani, P. Why it's quick to be square: modelling new and existing hierarchical menu designs. In *CHI '10*, ACM (2010), 1371–1380.

2. Anderson, J., and C., L. *The Atomic Components of Thought*. L. Erlbaum Associates Inc., 1998.

3. Anderson, J. R., Matessa, M., and Lebiere, C. Act-r: a theory of higher level cognition and its relation to visual attention. *Hum.-Comput. Interact. 12*, 4 (Dec. 1997), 439–462.

4. Bailly, G., Lecolinet, E., and Nigay, L. Flower menus: a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In *AVI '08*, ACM (2008), 15–22.

5. Bailly, G., Oulasvirta, A., Timo, K., and Hoppe, S. Menuoptimizer: Interactive optimization of menu systems. In *UIST'13, ACM. To appear* (2013).

6. Brumby, D. P., and Howes, A. Good enough but I'll just check: web-page search as attentional refocusing. In *Proceedings of the Sixth International Conference on Cognitive Modeling* (2004), 46–51.

7. Brumby, D. P., and Howes, A. Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *HumanComputer Interaction 23*, 1 (2008), 1–46.

8. Byrne, M. Act-r/pm and menu selection: Applying a cognitive architecture to hci. IJHCS 55 (2001), 41–84.

9. Byrne, M. D., Anderson, J. R., Douglass, S., and Matessa, M. Eye tracking the visual search of click-down menus. In *CHI '99*, ACM (1999), 402–409.

10. Card, S. K. User perceptual mechanisms in the search of computer command menus. In *CHI '82*, ACM (New York, NY, USA, 1982), 190–196.

11. Card, S. K., P., T., and Newell, A. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., 1983.

12. Cockburn, A., and Gutwin, C. A predictive model of human performance with scrolling and hierarchical lists. *HumanComputer Interaction 24*, 3 (2009), 273–314.

13. Cockburn, A., and Gutwin, C. A model of novice and expert navigation performance in constrained-input interfaces. *ACM Trans. Comput.-Hum. Interact. 17*, 3 (July 2010), 13:1–13:38.

14. Cockburn, A., Gutwin, C., and Greenberg, S. A predictive model of menu performance. In *CHI '07*, ACM (2007), 627–636.

15. Cox, A. L., and Silva, M. M. The role of mouse movements in interactive search. In *Annual Meeting of the Cognitive Science Society*, Lawrence Erlbaum (2006), 1156–1161.

16. Fu, W.-T., and Pirolli, P. Snif-act: a cognitive model of user navigation on the world wide web. *Hum.-Comput. Interact. 22*, 4 (Nov. 2007), 355–412.

17. Halverson, T., and Hornof, A. J. Local density guides visual search: Sparse groups are first and faster. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (2004).

18. Halverson, T., and Hornof, A. J. The effects of semantic grouping on visual search. In *CHI EA '08*, ACM (N, 2008), 3471–3476.

19. Halverson, T., and Hornof, A. J. A computational model of active vision for visual search in humancomputer interaction. *HumanComputer Interaction 26*, 4 (2011), 285–314.

20. Hornof, A. J. Cognitive strategies for the visual search of hierarchical computer displays. *Hum.-Comput. Interact. 19*, 3 (Sept. 2004), 183–223.

21. Hornof, A. J., and Kieras, D. E. Cognitive modeling reveals menu search in both random and systematic. In *CHI '97*, ACM (N, 1997), 107–114.

22. Hornof, A. J., and Kieras, D. E. Cognitive modeling demonstrates how people use anticipated location knowledge of menu items. In *CHI '99*, ACM (1999), 410–417.

23. Howes, A., Lewis, R. L., and Vera, A. Rational adaptation under task and processing constraints: implications for testing theories of cognition and action. *Psychological Review 116*, 4 (2009), 717–751.

24. Jacobs, A. M., and Grainger, J. Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human perception and performance 20*, 6 (1994), 1311.

25. John, B. E., Prevas, K., Salvucci, D. D., and Koedinger, K. Predictive human performance modeling made easy. In *CHI '04*, CHI '04, ACM (N, 2004), 455–462.

26. Kaptelinin, V. Item recognition in menu selection: the effect of practice. In *INTERACT '93 and CHI '93*, CHI '93, ACM (1993), 183–184.

27. Kieras, D. E., and Meyer, D. E. An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Hum.-Comput. Interact. 12*, 4 (Dec. 1997), 391–438.

28. Lane, D. M., Napier, H. A., Batsell, R. R., and Naman, J. L. Predicting the skilled use of hierarchical menus with the keystroke-level model. *Hum.-Comput. Interact. 8*, 2 (June 1993), 185–192.

29. Lee, E., and McGregor, J. Minimizing user search time in menu retrieval systems. *Human Factors 27*, 2 (1985), 157–162.

30. MacKenzie, I. S. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact. 7*, 1 (Mar. 1992), 91–139.

31. Malacria, S., Bailly, G., Harrison, J., Cockburn, A., and Gutwin, C. Promoting hotkey use through rehearsal with exposehk. In *CHI '13*, ACM (2013), 573–582.

32. McDonald, J. E., Stone, J. D., and Liebelt, L. S. Searching for items in menus: The effects of organization and type of target. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 27*, 834 (Oct. 1983), 834–837.

33. Mehlenbacher, B., Duffy, T. M., and Palmer, J. Finding information on a menu: linking menu organization to the user's goals. *Hum.-Comput. Interact. 4*, 3 (Sept. 1989), 231–251.

34. Newell, A., and Rosenbloom, P. S. The soar papers (vol. 1). MIT Press, Cambridge, MA, USA, 1993, ch. Mechanisms of skill acquisition and the law of practice, 81–135.

35. Nilsen, E., and Evans, J. Exploring the divide between two unified theories of cognition: modeling visual attention in menu selection. In *CHI EA '99*, ACM (1999), 288–289.

36. Nilsen, E. L. *Perceptual-motor control in human-computer interaction*. PhD thesis, Ann Arbor, MI, USA, 1991. UMI Order No. GAX91-35659.

37. Norman, K. L. *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*. Greenwood Publishing Group Inc., Westport, CT, USA, 1991.

38. Sears, A., and Shneiderman, B. Split menus: effectively using selection frequency to organize menus. *ACM Trans. Comput.-Hum. Interact. 1*, 1 (Mar. 1994), 27–51.

39. Smith, M. R., Lewis, R. L., Howes, A., Chu, A., and Green, C. More than 8,192 Ways to Skin a Cat : Modeling Behavior in Multidimensional Strategy Spaces. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, B. C. Love, K. McRae, and V. M. Sloutsky, Eds., Cognitive Science Society (Austin, Tx, 2008), 1441–1446.

40. Somberg, B. L. A comparison of rule-based and positionally constant arrangements of computer menu items. In *CHI '87*, ACM (1987), 255–260.