

An Integrated Model of Eye Movements and Visual Encoding

Dario D. Salvucci
Cambridge Basic Research

Please address correspondence to:

Dario Salvucci
Cambridge Basic Research
Four Cambridge Center
Cambridge, MA 02142
1-617-374-9669
dario@cbr.com

Running head:

Eye Movements and Visual Encoding

In Press, *Cognitive Systems Research*

January 15, 2001

Abstract

Recent computational models of cognition have made good progress in accounting for the visual processes needed to encode external stimuli. However, these models typically incorporate simplified models of visual processing that assume a constant encoding time for all visual objects and do not distinguish between eye movements and shifts of attention. This paper presents a domain-independent computational model, EMMA, that provides a more rigorous account of eye movements and visual encoding and their interaction with a cognitive processor. The visual-encoding component of the model describes the effects of frequency and foveal eccentricity when encoding visual objects as internal representations. The eye-movement component describes the temporal and spatial characteristics of eye movements as they arise from shifts of visual attention. When integrated with a cognitive model, EMMA generates quantitative predictions concerning when and where the eyes move, thus serving to relate higher-level cognitive processes and attention shifts with lower-level eye-movement behavior. The paper evaluates EMMA in three illustrative domains — equation solving, reading, and visual search — and demonstrates how the model accounts for aspects of behavior that simpler models of cognitive and visual processing fail to explain.

Keywords: eye movements, cognitive models, reading, visual search, computational models.

Introduction

Models of cognition have recently taken great strides in interacting with the visual world. Early models of cognitive processing often assumed that visual stimuli were already encoded and stored as internal representations. Recently, however, many models have begun also to account for the visual processing needed to encode stimuli through the planned movement of visual attention; for instance, researchers have developed such models in domains including reading (e.g., Just & Carpenter, 1980), arithmetic (e.g., Suppes, 1990), analogy (e.g., Salvucci & Anderson, in press), and human-computer interfaces (e.g., Anderson, Matessa, & Lebiere, 1997). These models specify how cognition directs visual attention to various parts of the visual stimuli, encodes these parts into internal representations, and synthesizes the encoded representations into higher-level representations and abstractions. The ability to account for both cognitive and visual processing provides the models with a higher degree of realism and plausibility and enables them to interact with the visual world through running computer simulations (e.g., Byrne & Anderson, 1998; Kieras & Meyer, 1997).

While these models of cognition have made strides in accounting for visual processing, most still emphasize cognitive aspects of behavior and thus take a somewhat simplistic view of eye movements and visual encoding (with notable exceptions, as discussed later). In particular, the models often make two fundamental assumptions either explicitly or implicitly. First, many models assume a direct correspondence between unobservable attention shifts and observable eye movements — that is, where people focus their attention is the same as where they look (e.g., Salvucci & Anderson, in press). Second, many models assume that the encoding of all visual objects requires the same amount of processing time — that is, what people encode does not affect how long they need to encode it (e.g., Anderson, Matessa, & Lebiere, 1997). Researchers agree that these assumptions, while they may hold in some cases, do not hold in general (see Henderson, 1992; Kowler, 1990; Rayner, 1995). As will be discussed further, these assumptions

seriously limit the ability of such models to account for visual processing behavior: while the models can account for behavior in terms of shifts of attention from one stimulus object to another, they cannot rigorously account for behavior in terms of actual eye movements from one location to another.

This paper presents a computational model, EMMA, that serves as a bridge between observable eye movements and the unobservable cognitive processes and shifts of attention that produce them. EMMA¹ (Eye Movements and Movement of Attention) incorporates a close but indirect link between eye movements and visual encoding, enabling the model to account for common eye-movement phenomena such as multiple fixations or “skipped” fixations (i.e., when a visual object is not fixated directly but rather encoded peripherally). EMMA takes a minimalist approach in attempting to describe these two processes as simply as possible. Concerning eye movements, the model describes whether or not eye movements occur, when they occur, and where they land with respect to their targets. Concerning visual encoding, the model describes how peripheral viewing and object frequency affect the time needed to encode a visual object into an internal representation. EMMA is not intended to capture all known phenomena in the vast literature on eye movements and visual attention and encoding. Instead, it is intended to demonstrate how a minimal description of these processes can significantly facilitate the modeling of cognitive and visual processing for common domains. In addition, because it produces quantitative predictions about when and where the eyes move, EMMA serves as a useful tool for generating behavior at a fine-grained level and for comparing predicted behavior to observable eye-movement data.

In addition to specifying the processes of eye movements and visual encoding, EMMA also describes the interface between these two processes and a central cognitive process. Like the processes themselves, the interface is quite minimal: when cognition requests an encoding of some visual object (and thus a shift of visual attention), EMMA handles the encoding process and resulting eye movements and provides the visual object to cognition when encoding is complete. This minimal interface allows EMMA to be easily integrated with various types of

cognitive processors. This paper describes the integration of EMMA with cognitive models developed in the ACT-R cognitive architecture (Anderson & Lebiere, 1998). However, it is important to note that EMMA can be integrated with any cognitive model or architecture that predicts shifts of visual attention from one visual object to another.

The EMMA model draws its inspiration from models of eye-movement control in reading, particularly the E-Z Reader model of Reichle, Pollatsek, Fisher, and Rayner (1998). Reading is a domain in which researchers have rigorously studied (and debated) the connection between higher-level cognitive processing and lower-level eye-movement behavior (e.g., O'Regan, 1990, 1992; Rayner, 1995, 1998; Rayner & Morris, 1990). (See also more general discussions in Kowler, 1990, and Viviani, 1990.) Models of reading have succeeded in accounting for both cognitive aspects of behavior, such as effects of word frequency and predictability on gaze durations (e.g., Just & Carpenter, 1980; Reichle et al., 1998), and oculomotor aspects of behavior, such as distributions of word landing positions (e.g., O'Regan, 1992; Reichle, Rayner, & Pollatsek, 1999; Reilly & O'Regan, 1998). These models are intended specifically for the domain of reading, and although they offer numerous insights into eye-movement control in other contexts, the generalization of the models to other domains is not obvious and has not yet been addressed. EMMA represents an attempt to extract the important principles from these models and generalize them into a domain-independent model of eye-movement control.

To validate the EMMA model, this paper describes applications to three domains: equation solving, reading, and visual search. All three domains involve interesting aspects of eye-movement behavior that simple models of eye movements and visual encoding cannot account for, such as skipped fixations resulting from peripheral encoding and gaze-duration effects resulting from different visual object frequencies. The exposition describes models of each domain that utilize EMMA to model eye movements and visual encoding, and ACT-R (Anderson & Lebiere, 1998) to model cognitive processing. These models demonstrate how EMMA allows models of higher-level cognition to account for lower-level effects of eye movements and encoding. While these applications only begin to cover the wide expanse of

possible domains, they do provide a good illustration of how a single general model can account for visual-processing behavior across different cognitive tasks.

Two Common Assumptions for Modeling Visual Processing

As mentioned, many models of cognitive and visual processing (e.g., Anderson, Matessa, & Lebiere, 1997; Byrne & Anderson, 1998; Card, Moran, & Newell, 1983; Ehret, 1999; Just & Carpenter, 1980; Lohse, 1993; Salvucci & Anderson, in press; Suppes, 1990; Suppes et al., 1983; Thibadeau, Just, & Carpenter, 1982) make at least one of two simplifying assumptions concerning eye movements and visual encoding — namely, that encoding time is constant for all visual objects and that eye movements directly correspond to focused attention. These assumptions have a serious impact on the development and evaluation of such models with observable data. Let us illustrate this issue with two tasks studied in this paper. Figure 1 shows sample screens and eye-movement protocols taken from the equation-solving and reading tasks. In the equation-solving task, students solved equations of a particular form by encoding each of four values in the equation and performing several straightforward computations (details are discussed later); after some practice, students became very efficient at this and could solve an equation in only a few seconds. In the reading task, students simply read medium-length sentences and were tested occasionally for comprehension. The plotted points on the figures represent students' eye movements during a single trial; larger points represent fixations and lighter points represent later samples.

[[Insert Figure 1 approximately here.]]

The first simplifying assumption is that eye movements directly correspond to visual attention, such that the models account only for shifts of focused attention between visual objects and not the actual eye movements themselves. This distinction between attention shifts and eye movements may seem like a subtle and unimportant one, but in fact the distinction is critical to developing models with observable data. In equation solving, Figure 1(a) illustrates

how students sometimes encode the outermost values in their peripheral vision, and thus eye-movement records show no apparent fixations on these values. Figure 1(b) illustrates how students sometimes produce multiple fixations on a single value (e.g., fixations 2-3-4 and 7-8) after undershooting or overshooting the value with an eye movement. In reading, Figure 1(c) illustrates how readers need not fixate every word as they read, even as they maintain almost perfect comprehension of the sentence (e.g., Schilling, Rayner, & Chumbley, 1998). In addition, research has shown that people begin processing words before their eyes reach the word by previewing the word in parafoveal or peripheral vision (McConkie & Rayner, 1975, 1976). For both tasks, a simple model that attends to each visual object once but does not dissociate eye movements and visual attention cannot account for these (commonly observed) phenomena.

The second simplifying assumption is that encoding of all visual objects into an internal representation requires the same amount of processing time. While researchers have found a number of factors that influence encoding time, this paper focuses on two key factors: the frequency with which the object is encoded; and the object's eccentricity, or distance from the current position of the eye's gaze. In reading, readers show a tendency to skip high-frequency words, and when they do fixate words, they spend less time processing high-frequency words (e.g., Schilling, Rayner, & Chumbley, 1998). Readers also require more time to encode words with greater eccentricity — that is, words that are farther in the periphery (Rayner & Morrison, 1981). In equation solving, students are more likely to skip outer values due to the fact that these values are high-frequency (one-digit) numbers. Again, a simple model that requires equal encoding times for all objects regardless of frequency and eccentricity cannot account for these results.

The EMMA Model

EMMA eliminates these two major assumptions and provides a formal computational model that helps to account for many of the above phenomena. The equation-solving and reading examples show how, with only a simple model of eye movements and visual encoding, even

accurate models of cognitive and visual processing for a task can produce predictions that differ significantly from observed eye-movement data. This problem can make it very difficult for researchers to develop initial prototype models through exploratory data analysis, evaluate models by comparing model predictions to observed data, and refine models by examining the mismatches between model predictions and observed data. By bridging the gap between cognitive processes, attention shifts, and eye movements, EMMA greatly facilitates the use of eye-movement data for modeling cognition and for understanding visual processing behavior.

As a minimal model, EMMA clearly does not address all known phenomena related to eye movements and visual encoding. For instance, it does not address how the eyes move within a large or complex visual object to encode different regions of the object, as might be necessary for image scanning and recognition (e.g., Noton & Stark, 1971). Also, it accounts only for normal voluntary saccades and does not address other types of eye movements such as express saccades or smooth movements (see Fischer, 1992; Kowler, 1990). Instead, this paper focuses on the effects of two important factors — object frequency and foveal eccentricity — on the interaction between voluntary eye movements, visual encoding, and cognitive processes. In particular, the paper demonstrates how a minimal model incorporating these factors produces quantitative predictions about observable eye movements that nicely capture behavior in three common domains.

EMMA

The EMMA model can be described in three basic components. First, it specifies the factors involved in attending to and encoding a visual object into an internal representation. Second, the model specifies the temporal and spatial aspects of eye movements as they arise from shifts of attention. Third, it specifies the control flow of the interactions between eye movements, visual encoding, and a cognitive processor. Each component is now described in turn.

Visual Encoding

EMMA uses a “spotlight” metaphor of visual attention that selects a single region of the visual field for processing (see Anderson, Matessa, & Lebiere, 1998; Cave & Bichot, 1999). When cognition requests a shift of attention to a new visual object, EMMA encodes the visual object into an internal representation. The time T_{enc} needed to encode object i is computed as follows:

$$T_{enc} = K \left[-\log f_i \right] e^{k \cdot i}$$

The parameter f_i represents the frequency of the object being encoded, specified as a normalized value in the range (0,1). The parameter i represents the eccentricity of the object, measured as the distance from the current eye position to the object in units of visual angle. Thus, encoding time increases as object eccentricity increases and as object frequency decreases. The constants K and k scale the encoding time and the exponent, respectively. To reflect variability in the system, the model assumes that encoding time is distributed as a gamma distribution with mean T_{enc} and standard deviation equal to one-third the mean (like Reichle et al., 1998).

This characterization of encoding time is based on empirically observed phenomena, particularly in the domain of reading. Research has shown that readers spend less time looking at frequent words than infrequent words (e.g., Schilling, Rayner, & Chumbley, 1998). In addition, their time spent looking at a word is inversely related to the log of the word’s frequency (Just & Carpenter, 1980, 1984). Additional work has shown that words farther in the periphery — that is, with greater eccentricity — require more time to encode (Rayner & Morrison, 1981).

Eye Movements

In EMMA, a shift of attention initiates an eye-movement program to the attended object. The eye movement runs through two stages, *preparation* and *execution* (cf. Byrne & Anderson, 1998; Kieras & Meyer, 1997). Preparation represents the retractable, or “labile” (Reichle et al.,

1998), stage of the eye-movement program — that is, it can be canceled if cognition requests another preparation before the first terminates. Execution includes both non-retractable, or “non-labile”, programming, which cannot be canceled, and the actual eye movement itself. The motivation for a separation of labile and non-labile programming comes from results showing that eye movements can be canceled for only for a certain time threshold after the initiation of saccadic programming (Becker & Jürgens, 1979). The completion of a saccade initiates preparation for a new saccade to the same visual object in the event that encoding has not yet completed.

EMMA describes both the temporal and spatial characteristics of eye movements. With respect to temporal characteristics, the model includes two parameters T_{prep} and T_{exec} that describe the time required for preparation and execution, respectively. Preparation time T_{prep} is estimated to be 135 ms as described shortly. Execution time T_{exec} includes 50 ms for non-labile programming (cf. Becker & Jürgens, 1979), 20 ms for saccade execution, plus an additional 2 ms for each degree of visual angle subtended by the saccade (Fuchs, 1971). The total time to prepare and execute a saccade closely resembles saccade latencies of approximately 200 ms cited in many previous studies (e.g., Anderson, Matessa, & Lebiere, 1997; Fuchs, 1971; Russo, 1978). Again, the model adds variability by assuming that T_{prep} and T_{exec} are distributed as a gamma distribution with a standard deviation equal to one-third the mean.

With respect to spatial characteristics, EMMA provides a simple formalization of where an eye movement lands with respect to its desired destination. Given a saccade to a particular object, the model assumes that the landing point follows a Gaussian distribution around the center of the object. The distribution is given a standard deviation of 0.1 times the distance from saccade origin to intended destination as has been estimated empirically (see interpretation of Timberlake, Wyman, Skavenski, & Steinman, 1972, by Kowler, 1990). Some researchers have posited an “undershoot bias” such that saccades are more likely to undershoot rather than overshoot a desired target (e.g., Becker & Fuchs, 1969; Henson, 1978). While this bias could be

incorporated into the model by skewing or otherwise altering the distribution, EMMA opts for the simpler unbiased model in line with its minimalist approach.

Control Flow

The control flow of the EMMA model describes how cognition, visual encoding, and eye movements interact as interdependent processes. When cognition requests an attention shift to a new visual object, EMMA begins encoding the object while an eye movement is prepared and (possibly) executed. Depending on the order in which the processes complete, various scenarios arise in their interaction. This section describes the various possible scenarios, illustrated in Figure 2 as the interaction of the cognitive processor (Cognition), visual encoding (Vision), eye-movement preparation (Eye-Prep), and eye-movement execution (Eye-Exec).

[[Insert Figure 2 approximately here.]]

In the simplest case, encoding requires the same amount of time as an eye movement. In this case, the visual-encoding module works on encoding the object while the eye-movement module runs through each of its two stages. Figure 2(a) illustrates this case, with horizontal bars showing the execution of each module and stage. Note that the cognitive processor cedes control to visual encoding and receives control back when encoding is complete.

Another two cases arise when encoding completes and cognition requests a subsequent shift of attention before the original eye movement has completed. If the attentional shift occurs during eye-movement preparation, the eye movement is canceled and a new eye movement is begun, as shown in Figure 2(b). If the attentional shift occurs during eye-movement execution, execution continues to run to completion while a new eye movement is begun, as shown in Figure 2(c).

If the eye movement completes before encoding completes, encoding continues and a new eye-movement is prepared, as shown in Figure 2(d). However, because the eye movement has (presumably) brought the fovea nearer to the visual object, encoding speed increases accordingly.

The model calculates the remaining encoding time as follows: given the old total encoding time T_{enc} , the amount of completed processing time $T_{completed}$, and a new total encoding time T'_{enc} , the remaining encoding time is set to $(1 - (T_{completed} / T_{enc})) \cdot T'_{enc}$. As shown in the figure, the second eye-movement program eventually produces a refixation. This aspect of the model helps to account for behavior when objects are distant from the fovea and encoding time is very long, since the model can refixate objects to decrease eccentricity and facilitate encoding.

Two points should be noted with respect to EMMA's specification of control flow. First, as described earlier, EMMA incorporates an indirect link between encoding and eye movements that decouples unobservable attention shifts and observable eye movements. In particular, attention shifts occur at the initiation of encoding — that is, when cognition requests an attention shift, EMMA immediately shifts attention and begins to encode the desired object. In contrast, while eye-movement programming initiates at the same time as the attention shift, the actual eye movement only occurs after programming completes at some later time; thus, there is a temporal lag between unobservable attention shifts and observable eye movements. Second, as EMMA stands currently, encoding and cognition proceed normally during a saccadic eye movement. Some researchers have noted a phenomenon of “saccadic suppression” in which eye movements interrupt encoding and/or cognition (see Matin, 1974; Volkman, 1986); however, others have found that eye movements may not interrupt encoding and cognition in certain situations (see Irwin, 1998). Again in line with the minimalist approach, EMMA opts for the simpler model and allows encoding and cognition to proceed during an eye movement.

Parameter Settings

Because EMMA serves as a general model for all domains, its parameter values should hold across all possible models. Table 1 shows the list of EMMA parameters, whether they were preset or estimated, and their resulting values. As mentioned, T_{exec} was preset according to values found in previous studies: 50 ms for non-retractable saccade programming (Becker & Jürgens, 1979), and 20 ms for saccade execution plus 2 ms for each degree subtended by the

saccade (Fuchs, 1971). The remaining parameters were estimated to produce the best fits for the equation-solving and reading domains. (The visual-search study was performed later using these estimated values.) The estimated value of 135 ms for T_{prep} is approximately the same as the value of 150 ms used in the E-Z Reader model (Reichle et al., 1998).

[[Insert Table 1 approximately here.]]

Discussion

To place EMMA in the context of existing work, let us compare it to several related models. As mentioned, EMMA reflects a synthesis of ideas from a number of existing models of eye-movement control, particularly those that emphasize the influence of cognitive processes (e.g., Just & Carpenter, 1980; Morrison, 1984; Rayner, Reichle, & Pollatsek, 1998; Reichle et al., 1998; Thibadeau, Just, & Carpenter, 1982). Of these models, EMMA is most closely based on the E-Z Reader model of eye-movement control in reading (Rayner, Reichle, & Pollatsek, 1998; Reichle et al., 1998). While the details differ, EMMA and E-Z Reader share many similarities, including a two-stage eye-movement program and a frequency- and eccentricity-based computation of encoding time. However, there are two important differences between these models. First and most importantly, E-Z Reader is intended to model eye movements specifically in the domain of reading, whereas EMMA is intended to model eye movements in a domain-independent fashion. Second, the two models differ with respect to the timing of the encoding and eye-movement processes. In E-Z Reader, encoding and eye-movement programming begin at different times: completion of a “familiarity check” initiates eye-movement programming while completion of full processing initiates an attention shift and encoding. In EMMA, encoding and eye-movement programming begin simultaneously: both processes start when cognition requests an attention shift (like Morrison, 1984). To a large extent, this difference is one consequence of EMMA’s generality: while E-Z Reader knows the next saccade target (i.e., the next word), EMMA does not know the target until directed by cognition.

Another illustrative comparison arises between EMMA and the visual processing models built into two cognitive theories, ACT-R (Anderson & Lebiere, 1998) and EPIC (Kieras & Meyer, 1997). These theories incorporate both a sophisticated cognitive processor for modeling cognition, and also a number of perceptual-motor mechanisms for modeling the interaction of cognition with the outside world. ACT-R's perceptual mechanisms — both the former (the visual interface: Anderson, Matessa, & Lebiere, 1997) and the current (ACT-R/PM: Byrne & Anderson, 1998) — incorporate only a basic model of eye movements and visual encoding that makes both of the simplifying assumptions discussed in the introduction — namely, constant-time encoding and lack of dissociation of eye movements and visual attention. The EPIC model is more sophisticated than the ACT-R model but falls short of the more rigorous treatment in EMMA with respect to voluntary saccadic movement. For instance, EPIC provides a separation between eye movements and visual attention but requires that the cognitive processor request eye movements separately; they do not simply arise from attention shifts. Also, EPIC has only a discrete characterization of eccentricity and includes some special-case assumptions (e.g., “centering” movements to refixate objects) and free parameters that require adjustment for each new domain. EMMA builds upon this work by providing a minimalist but rigorous model with few estimated parameters and no task-specific mechanisms. In fairness, EPIC provides an account for smooth and involuntary movements that EMMA does not presently include.

For three domain applications, the work described here extends the ACT-R theory to incorporate EMMA and demonstrates how the integrated framework can account for many empirical phenomena that the basic ACT-R framework cannot. However, it is important to note that EMMA (or its basic principles) could be integrated with any cognitive processor. The cognitive processor could come from a general cognitive architecture (e.g., Soar: Laird, Newell, & Rosenbloom, 1987, Newell, 1990; EPIC: Kieras & Meyer, 1997; CAPS: Just & Carpenter, 1992); it could be a domain-specific model of cognition for a particular task, such as a cognitive model of reading (e.g., Just & Carpenter, 1980; Thibadeau, Just, & Carpenter, 1982) or arithmetic

(e.g., Suppes, 1990); or it could be any arbitrary cognitive processor that predicts shifts of attention from one visual object to another.

Illustrative Applications

This section applies and evaluates the EMMA model in three domains. The equation-solving domain emphasizes the model's predictions concerning peripheral encoding, spatial variability, and the interaction of eye movements with cognitive computation. The reading domain emphasizes the effects of word frequency on various fixation duration and fixation probability measures. The visual-search domain focuses on total latency and skipped fixations in a randomized menu selection task. All together, the domains provide three complementary evaluations of EMMA and demonstrate the model's ability to account for a number of aspects of visual-processing behavior in observable eye movements.

All three domain applications utilize ACT-R (Anderson & Lebiere, 1998) as a framework for developing a cognitive model that interacts with the eye-movement and visual encoding components of EMMA. ACT-R is a general theory of cognition based on condition-action production rules that has successfully modeled behavior in a wide range of domains, including arithmetic, choice, analogy, and scientific discovery (see Anderson & Lebiere, 1998). Because the details of ACT-R are peripheral to this exposition, readers are referred to Anderson and Lebiere (1998) for a thorough description. However, it is important to note three reasons for which ACT-R was chosen for the purposes of integration with EMMA. First, the theory provides many useful constraints on the structure and running of the cognitive model; these constraints force the model to obey various cognitive and perceptual-motor limitations, thus leading to more psychologically (and neurally) plausible models. Second, the theory is firmly grounded in a computational framework that incorporates perceptual-motor modules (Byrne & Anderson, 1998), facilitating implementation of EMMA as a straightforward addition to and adaptation of these modules. Third, because the theory has been applied to a variety of task domains, each of

these domains offers future potential to provide further validation of EMMA in addition to the domain applications presented here.

Equation Solving

The equation-solving task involves solving equations of a particular form, namely

$$a x / B = A / b ,$$

by computing $x = (A/a)(B/b)$; for instance, the solution to

$$4 x / 21 = 24 / 7$$

is $x = (24/4)(21/7) = 18$. (See Figures 1(a) and 1(b) for sample problems and eye-movement protocols.) After some practice, college undergraduates solve these problems accurately and quickly (in only a few seconds). Note that while most studies of equation solving focus on younger students with various types of equations, this study emphasizes the visual processing evident in practiced behavior when solving equations of a particular form.

The equation-solving task is interesting for at least two important reasons. First, as discussed in the introduction, students often do not follow a strict pattern of fixating each value once, but rather sometimes skip over values or refixate values. Second, the cognition required in computing the intermediate and final results affects students' observed eye movements. Thus, to accurately represent behavior in this task, a model must account for both lower-level phenomena present in students' eye movements and higher-level effects of cognition on these eye movements. The model and results described here build on results from an earlier model described in previous work (Salvucci, 2000).

Data

The data used in this study come from an earlier experiment designed to demonstrate and evaluate methods for interpreting eye-movement data (Salvucci & Anderson, 1998, in press-b).

In the experiment, students solved the equations described above using an *instructed-strategy* paradigm: during each of four sessions, students were instructed to execute a particular problem-solving strategy. The four strategies used for each of the four sessions are shown in Table 2. The strategies dictated when to encode each value — for instance, strategy $[a B A b]$ encodes values left-to-right whereas strategy $[a A B b]$ encodes values left-to-right in pairs. In addition, all strategies dictated when to compute intermediate results, namely as soon as possible — for instance, immediately after encoding A for either strategy $[a B A b]$ or $[a A B b]$. The instructed-strategy paradigm used in the experiment is critical to the current study because it greatly constrains the cognitive model to known strategies and allows us to focus on EMMA's predictions as they arise from these strategies. The data set includes protocols from four subjects and a total of 369 trials.

[[Insert Table 2 approximately here.]]

Model

The model of student behavior comprises standard ACT-R production rules that implement each of the four instructed strategies. The model takes a minimalist approach in the sense that it assumes the simplest possible rules that can implement each strategy; in essence, the model for each strategy simply encodes each value in the specified order, performs computation as soon as possible (as students were instructed), and types out the final result. The model has two parameters: *effort*, or the time needed for each production to fire, preset to 10 ms; and *strength*, or speed at which productions retrieve declarative chunks, estimated to be 1.0. In addition, the frequencies of one-digit and two-digit values were preset at .10 and .01, respectively.

Results

The following analysis of empirical results and model predictions focuses on student behavior with respect to visual processing. For the empirical results, because some trials include review of equation elements or other behavior not captured by the model, this analysis is limited to protocols with four or fewer gazes on the equation values (60% of all trials). The model predictions include results from 400 model simulations evenly divided among the four strategies. The following analysis aggregates consecutive fixations on the same object into gazes as described by Just and Carpenter (1984).

Figure 3 illustrates the students' and model's use of peripheral encoding as the percentage of protocols with four, three, and two gazes on the equation values. Students looked at all four values in only about two-thirds of all trials: they exhibited four gazes in 69% of the trials, three gazes 29%, and two gazes 2%. Because EMMA can produce skipped gazes by canceling saccades when encoding occurs quickly (i.e., before eye-movement preparation is complete), it also exhibits peripheral encoding of values. The model's predictions closely fit the observed data, $R > .99$. In addition, the model refixates (i.e., produces multiple fixations on) the equation values 19% of the time, compared to 30% for students.

[[Insert Figure 3 approximately here.]]

Figure 4 shows the effects of performing computation while looking at the equation values. Because students computed intermediate results (i.e., A/a and B/b) as soon as possible (as instructed), gazes are classified based on the number of computations performed during the gaze: no computation (NC), one computation (1C), and two computations (2C). For instance, for the strategy $[a A B b]$, the a and B gazes would be classified as NC, the A gaze as 1C (A/a is computed), and the b gaze as 2C (B/b and x are computed). As the figure shows, students spend approximately an additional 300 ms for each computation performed during a gaze; these differences are very significant, $p < .001$. The model computes results during 1C and 2C gazes by

retrieving a declarative chunk that represents the solution to the computation. The time for this retrieval allows the model to capture the observed effect of computation, $R > .99$.

[[Insert Figure 4 approximately here.]]

Figure 5 manifests the effects of skipping on gaze durations for NC and 1C gazes. It shows the mean durations for these gazes when the next value in the instructed strategy is skipped (1S) and when it is not skipped (NS). When students skipped the next value, their eyes remained on the current value while encoding the next value peripherally, thus increasing the duration of the gaze on the current value. The effects of skipping, computation, and their interaction are all significant, $p < .05$. Because the model also maintains its gaze on the current value when encoding the next value, it captures the main effects of both skipping and computation. However, the model does not predict the interaction in the high value for the 1C-1S gazes; the problem seems to arise from the fact that the cognitive model does not adequately capture behavior for the various strategies rather than problems with EMMA itself. Nevertheless, the model fit remains good, $R = .92$.

[[Insert Figure 5 approximately here.]]

Figure 6 shows the spatial variability of the empirical data and the model results. The figure illustrates spatial variability along two axes: (a) the axis parallel to the vector from the saccade start point to the intended visual target, and (b) the axis perpendicular to this vector. For both axes, the empirical data and the model predictions exhibit a roughly Gaussian distribution centered around the visual target with similar standard deviations, $R > .97$ for both axes. As an aside, the equation-solving data do not seem to manifest a bias for undershooting the target on the parallel axis as has been observed in some studies (e.g., Becker & Fuchs, 1969; Henson, 1978).

[[Insert Figure 6 approximately here.]]

Reading

Of the many domains in which eye movements have been studied, reading has arguably received the most attention and in-depth analysis in the literature. Models of reading behavior have focused on eye movements to gain insight into reading's underlying processes (see Rayner, 1998). In particular, several researchers have proposed models of eye-movement control that specify where and when the eyes move during reading (e.g., Legge, Klitz, & Tjan, 1997; Morrison, 1984; O'Regan, 1992; Rayner, Reichle, & Pollatsek, 1998; Reichle et al., 1998; Reilly & O'Regan, 1998; Suppes, 1990; Thibadeau, Just, & Carpenter, 1982). These models do not attempt a full description of the syntactic and semantic processing necessary for reading, instead focusing on accounting for visual processing in terms of observed eye movements. However, some models do incorporate aspects of higher-level cognitive processing (e.g., lexical access) and demonstrate how these aspects can affect resulting eye movements (e.g., Just & Carpenter, 1980; Morrison, 1984; Reichle et al., 1998; Thibadeau, Just, & Carpenter, 1982).

This section describes an EMMA-based model of reading that, like the above models, emphasizes eye-movement control in accounting for where and when the eyes move. It also incorporates some minimal aspects of cognitive processing in its two-stage processing of words. First, it encodes the visual word stimulus into an internal representation that embodies the visual image. The time needed for this encoding process depends on word frequencies as dictated by EMMA. Second, it performs lexical access in that it retrieves the semantic representation that corresponds to the encoded visual representation. The current model currently incorporates no interesting predictions with respect to this process. However, in its use of a general cognitive architecture (ACT-R), it is quite feasible that extensions of the model could incorporate further syntactic and semantic processing and account for many higher-level cognitive aspects of reading behavior.

To evaluate the model as a model of eye-movement control, this work compares model predictions to the same empirical results used by Reichle et al. (1998) to evaluate their E-Z

Reader model. As mentioned, EMMA and E-Z Reader are closely related in their specifications of the interaction between eye movements and visual encoding. On the one hand, E-Z Reader can account for aspects of reading behavior in particular that EMMA cannot currently account for; for instance, E-Z Reader can accurately predict the distributions of fixation locations within words (Reichle, Rayner, & Pollatsek, 1999). On the other hand, EMMA serves as a domain-independent model applicable to arbitrary domains. Thus, the purpose of this study is to ensure that EMMA retains the essential characteristics of E-Z Reader and can account for eye movements in reading in addition to the other illustrative applications.

Data

Reichle et al. (1998) compiled their results from data collected in an experiment by Schilling, Rayner, and Chumbley (1998). In the experiment, students read 48 sentences of 8-14 words as their eye movements were recorded. Students were occasionally tested to ensure comprehension. The original experiment data set comprised data from 48 students reading each sentence once. The results compiled by Reichle et al. and presented here include only those data from trials with no interword regressions (36% of all trials).

Model

As for the equation-solving model, the EMMA reading model is developed using a minimalist approach. The model has five basic productions: move attention to the next word for encoding, retrieve the encoded word's visual representation, retrieve the semantic object chunk that corresponds to this visual representation, create a word if none exists, and terminate reading when completed. The reading model has the same two parameters as the equation-solving model: *effort* was again preset to 10 ms, and *strength* was preset to 5.0. Frequency parameters for each word were set and normalized according to the word's natural frequency of occurrence — namely, its frequency per one million words (Francis & Ku[^]cera, 1982) divided by one million.

Results

The analyses here follow Reichle et al.'s analysis of the effects of word frequency with respect to six fixation time and probability measures. First, the words were divided into classes 1-5 according to the log frequency by which they appear in standard text (Francis & Ku[^]cera, 1982); class 5 words were the most frequent (e.g., *the*), and class 1 words were the least frequent (e.g., *burglar*). The mean frequencies per one million words for each class are included in Table 3. Next, fixation times were analyzed using three measures: *gaze duration*, or summed duration of consecutive fixations on a word; *first-fixation duration*, or duration of only the first fixation on a word; and *single-fixation duration*, or fixation duration on a word fixated exactly once. Fixation probabilities were analyzed by computing the probabilities of fixating a word zero, one, or two times; these measures were termed *skip*, *one-fixation*, and *two-fixation probabilities*, respectively.

Table 3 shows the empirical data and model results for the three categories of durations, namely gaze, first-fixation, and single-fixation durations. The model predictions come from 10 simulations of each of 48 sentences (480 simulations total). For both data and model, all three measures show a strong trend in which high-frequency words (with a large class number) have shorter durations than low-frequency words. The model predictions reproduce this trend through EMMA's characterization of encoding time. The model provides an excellent fit to the observed gaze durations and reasonably good fits for the other two measures, $R = .93$ for all cases.

[[Insert Table 3 approximately here.]]

Table 4 shows the data and predictions for probabilities of producing zero, one, or two fixations on a word. Again there is a strong effect of word class: readers skip high-frequency words more often and refixate them less often than low-frequency words. Because the model predicts longer fixation durations on low-frequency words, it is more likely to fixate (and refixate) these words. Thus, although the quantitative fit of the model predictions contains some discrepancies, the model does capture the major trends in the empirical data, $R = .96$ for all cases.

The total root-mean-squared error over all duration and probability measures, computed as described by Reichle et al. (1998), is .362. This value is somewhat higher than the total root-mean-squared error of .198 reported for E-Z Reader 5 (Reichle et al., 1998) and .218 reported for E-Z Reader 6 (Reichle, Rayner, & Pollatsek, 1999). Nevertheless, EMMA's fit to the data can be considered a success given its generality in accounting for visual processing in reading as well as other domains.

[[Insert Table 4 approximately here.]]

Visual Search

The third domain application involves visual search in a menu-selection task (Byrne et al., 1999, based on Nilsen, 1991). In the task users were presented with a menu bar that specified a desired target, namely a number or letter. Users read the menu bar to encode the desired target and, using the mouse, clicked on the menu bar to reveal a vertical menu list of elements. Figure 7 shows a sample menu after the user has clicked on the menu; the menu bar in the upper portion specifies the desired target, and the menu list in the lower portion comprises a list of (in this case) six elements. Users then searched this list for the target and clicked on the target to end the trial. The menu items were presented in a random order, and thus the task essentially involved visual search of items and not memorization of item positions.

[[Insert Figure 7 approximately here.]]

Two groups of researchers have proposed cognitive models of behavior in this task. The model of Anderson, Matessa, and Lebiere (1997) assumes a top-down search strategy in which users start at the first element and encode each element in order until the target is found. The model of Hornof and Kieras (1997) employs either the top-down strategy or a random-search strategy; the random-search strategy search assumes that users start at any point on the list, evaluate whether this element is the target, and if not continue to the next random element. Both

groups evaluated their models according to trial latencies — that is, the time between the first mouse click on the menu bar and the second click on the target. To evaluate the models more rigorously, Byrne et al. (1999) examined user eye movements in the task and found that neither model captured behavior well at this level of detail. In particular, they found that users' eye movements were characteristically top-down and inconsistent with the predictions of a random-search strategy. However, the eye movements were also not strictly top-down in the sense that users did not always start with the first element and often skipped elements on the way down.

This section presents a simple model of behavior in the menu-selection task that embodies the top-down strategy but utilizes EMMA to make predictions about the eye movements resulting from this strategy. In particular, EMMA provides an explanation about how and when skipped fixations occur, thus helping to account for the phenomena observed by Byrne et al. (1999). Like the other illustrative models, this model takes a minimal approach in assuming a top-down strategy and does not account for more subtle effects in the data, such as when users skip well past the target. Nevertheless, the model demonstrates how, even with a simple cognitive model, EMMA helps models to account for observable eye-movement data.

Data

The data used here come from Byrne et al.'s (1999) results of their menu-selection experiment based on earlier experiments by Nilsen (1991). They collected both eye-movement and mouse-movement data for menus of six, nine, or twelve elements. The experiment included several varied factors: whether the target was a number or letter; whether the other menu elements were numbers or letters; and whether the target did or did not appear in the menu. Byrne et al. do not report the results of these factors and thus they are not addressed here. The data set includes data from 108 trials from each of 11 users for a total of 1188 trials.

Model

The ACT-R cognitive model for the task embodies the top-down strategy. It contains four search productions that attend to the next item, encode its value, terminate search if the item matches the target, and continue the search otherwise. It also contains three additional productions that move the mouse to the target item and click on it to end the trial. The target value is assumed to be known at the start of the trial, since trial time is computed beginning after users encoded the target and clicked on the menu bar to reveal the menu.

The menu-selection model utilizes the same EMMA parameter values as the other two models. For the sake of parsimony, it also uses the same values for *effort* and *strength* used in the reading model. It assumes a frequency value of 0.1 for numbers, as in the equation-solving model, as well as letters. The only other relevant parameters to the model are the parameters of the motor behavior needed to move and click the mouse. Because the model utilizes the standard motor modules built into ACT-R (Byrne & Anderson, 1998), these parameter values are simply the default values as determined in psychophysical empirical work and previous modeling efforts (see Byrne & Anderson, 1998); for instance, it uses Fitts' law to compute the time needed to move the mouse from one position to the next. Thus, all parameters of the model are constrained by previous settings and no new values were estimated.

Results

Figure 8 shows the mean trial times for both data and model, where model results are based on 100 simulation runs for each possible menu length and target position. The figure graphs trial times according to the position of the target for the six, nine, and twelve-item menus. The data show a clear linear trend with an increase of approximately 100 ms per target position. The model captures this trend through its top-down processing, requiring additional time to encode each item and compare it to the desired item, $R=.91$. Neither the data nor model exhibit any significant differences for the various length menus. The model fails to capture a subtler effect in the trial-time data, namely the slightly higher-than-expected times for positions 1 and 2.

[[Insert Figure 8 approximately here.]]

Figure 9 shows the ratio of first fixations on each menu position. The data strongly support a top-down search strategy, with almost no first fixations after position 5. However, they also indicate that users do not always start with the first item, but rather start at the initial items with decreasing frequency. Although the model always attends to and encodes the items in strict sequence, it sometimes does not fixate these items — that is, it sometimes encodes them quickly and cancels the eye-movement program before completing the movement. Thus, the model exhibits a decreasing likelihood of first fixations and provides a good fit to the data, $R=.92$. Although the model simulations for 12-item menus slightly overpredicts the likelihood of first fixations on position 2 versus position 1, overall the model, like the data, exhibits no systematic effect of menu length.

[[Insert Figure 9 approximately here.]]

As stronger support for the top-down strategy, Figure 10 shows the average number of fixations on each item position according to target position on a 9-item menu. The data, though somewhat noisy, show a general trend: a moderate number of fixations before the target, the most fixations on the target, and few fixations after the target. Again the model captures the general trend, $R=.69$. The model tends to overpredict fixations on the target given that it generally fixates the target while it moves the mouse there. Perhaps the most important deviation between model and data arises in the underprediction of fixations for items soon after the target, especially with the target at positions 1 and 2. This deviation can be attributed primarily to the fact that the simple cognitive model described here does not account for “passing over” the target even after it is encoded, as some users are reported doing (Byrne et al., 1999). However, it is interesting to note that EMMA does predict some fixations on the item immediately after the target due to overshooting saccades to the target.

[[Insert Figure 10 approximately here.]]

Summary

The three illustrative applications demonstrate how EMMA can produce good qualitative and quantitative fits to data across several domains. A simple model of eye movements and visual encoding could not account for almost any of the results reported here. For instance, let us consider the predictions of the presented ACT-R models without EMMA. As mentioned, ACT-R includes basic perceptual-motor mechanisms that assume a constant encoding time and do not distinguish eye movements from visual attention (Anderson & Lebiere, 1998; Byrne & Anderson, 1998). Thus, the ACT-R models could not account for the peripheral encoding and spatial variability in the equation-solving domain, the frequency effects on fixation duration and probability measures in the reading domain, and the peripheral encoding and first-fixation distributions in the visual-search domain. By incorporating the predictions of EMMA, the models can immediately account for all these phenomena.

General Discussion

This paper describes the EMMA model of eye movements and visual encoding and demonstrates how the model can help account for behavior in several illustrative domain applications. EMMA serves both a theoretical and practical role. As a theory, EMMA provides a formalization of the relationship between eye movements and attention shifts as directed by a cognitive processor. Using minimal mechanisms and parameters, the model helps to explain common eye-movement phenomena such as skipped fixations, multiple (or re-) fixations, and spatial variability in fixation locations. And by providing general domain-independent mechanisms, EMMA helps to generalize eye-movement models of specific domains such as reading into a broader account of eye movements across domains.

As a practical tool, EMMA provides rigorous quantitative predictions that can be directly compared to observable eye-movement data. While this paper demonstrates EMMA in the context of the ACT-R theory, EMMA could be integrated with any arbitrary cognitive

processor — the processor need only predict shifts of attention, and from there EMMA predicts when and where the eyes move. As a related point, any existing cognitive models with visual processing can immediately make use of EMMA to link attention shifts and eye movements. The only necessary addition to existing models is the specification of frequency values as required by EMMA; such values can and should be systematically maintained as default values across domains. In addition to EMMA's usefulness in predicting eye movements, its parameters could be employed in models that interpret eye movements — that is, map observed eye movements back to the cognitive processes that produced them (Salvucci & Anderson, 1998, in press-b).

The benefits of integrating EMMA with larger theories of cognition, particularly *cognitive architectures*, warrants further emphasis. Cognitive architectures, such as ACT-R and others (e.g., Just & Carpenter, 1992; Kieras & Meyer, 1997; Laird, Newell, & Rosenbloom, 1987; Newell, 1990), provide a rigorous framework for implementing cognitive models. This framework imposes numerous cognitive parameters and constraints on the model, such as how memory chunks are stored and retrieved, how these chunks are learned and strengthened with practice, and how they decay when ignored. Likewise, EMMA imposes visual-processing parameters and constraints on the model, such as when visual objects are encoded and when the eyes move to these objects. By integrating EMMA with a cognitive architecture, any model developed within the integrated framework necessarily incorporates aspects of both lower-level oculomotor processes and higher-level cognitive processes.

The illustrative applications in this paper provide good demonstrates of the interaction between oculomotor and cognitive processes. The equation-solving model successfully fits lower-level behavior in terms of gaze durations and fixation landing points. However, the fixation duration measures are clearly influenced by cognitive processing — namely, the computation of intermediate results during problem solving. This aspect of the model helps to explain why some gazes are significantly longer than others (approximately 200-300 ms for gazes during which no computation occurs, versus 500-900 ms for gazes during which computation occurs). The menu-

selection model also fits lower-level behavior in terms of first-fixation distributions and fixation counts. However, execution of this task requires additional processing; in particular, it must evaluate each item against the target item (incurring additional cognitive cost per item) and must move and click the mouse to terminate the trial. The model's integration of visual processing, cognitive processing, and additional motor movement produces grounded predictions that can more readily be compared to other aspects of user behavior — for instance, how and when users move the mouse to the target.

The reading model provides several interesting avenues for exploration of the interaction between oculomotor and cognitive processes. Rayner and others (Rayner, 1995; Rayner & Morris, 1990; Reichle, Rayner, & Pollatsek, 1999) have advocated a view of reading that stresses the importance of both processes, and several models of reading to date have illustrated the benefits of such a view (e.g., Just & Carpenter, 1980; Reichle et al., 1998; Thibadeau, Just, & Carpenter, 1982). For the reading model presented here, the fact that it is implemented in a general cognitive architecture enables a wide variety of possible future extensions incorporating syntactic and semantic processing. Again, any such incorporation would make immediate predictions about observable eye movements; for instance, if the model paused to parse ambiguous phrases or to construct a final representation after completing a sentence, it would predict longer gaze durations over the appropriate words as empirical studies have observed (see Just & Carpenter, 1984).

References

- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439-462.
- Becker, W., & Fuchs, A. F. (1969). Further properties of the human saccadic system: Eye movements and correction saccades with and without visual fixation points. *Vision Research, 9*, 1247-1258.
- Becker, W., & Jürgens, R. (1979). An analysis of the saccadic system by means of a double-step stimuli. *Vision Research, 19*, 967-983.
- Byrne, M. D., Anderson, J. A., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. In *Human Factors in Computing Systems: CHI 99 Conference Proceedings* (pp. 402-409). New York: ACM Press.
- Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson & C. Lebiere (Eds.), *The Atomic Components of Thought* (pp. 167-200). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. *Psychological Bulletin & Review, 6*, 204-223.
- Ehret, B. D. (1999). Learning where to look: The acquisition of location knowledge in display-based interaction. Doctoral Dissertation, George Mason University.
- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 31-45). New York: Springer-Verlag.

- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Fuchs, A. F. (1971). The saccadic system. In P. Bach-y-Rita, C. C. Collins, & J. E. Hyde (Eds.), *The Control of Eye Movements* (pp. 343-362). New York: Academic Press.
- Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 260-283). New York: Springer-Verlag.
- Henson, D. B. (1978). Corrective saccades: Effects of altering visual feedback. *Vision Research*, 18, 63-67.
- Hornof, A. J., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. In *Human Factors in Computing Systems: CHI 97 Conference Proceedings* (pp. 107-114). New York: ACM Press.
- Irwin, D. E. (1998). Lexical processing during saccadic eye movements. *Cognitive Psychology*, 36, 1-27.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 151-182). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kieras, D. E., & Meyer, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104, 3-65.
- Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movement. In E. Kowler (Ed.), *Eye Movements and their Role in Visual and Cognitive Processes* (pp. 1-70). New York: Elsevier Science Publishing.

- Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, *33*, 1-64.
- Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review*, *104*, 524-553.
- Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, *8*, 353-388.
- Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*, 899-917.
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578-586.
- McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*, 365-368.
- Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 667-682.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Nilsen, E. L. (1991). Perceptual-motor control in human-computer interaction (Tech. Rep. No. 37). Ann Arbor: University of Michigan, Cognitive Science and Machine Intelligence Laboratory.
- Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*, 929-942.
- O'Regan, J. K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Eye Movements and their Role in Visual and Cognitive Processes* (pp. 395-453). New York: Elsevier Science Publishing.
- O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 333-354). New York: Springer-Verlag.

- Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye Movement Research: Mechanisms, Processes, and Applications* (pp. 3-21). New York: Elsevier Science Publishing.
- Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin*, *124*, 372-422.
- Rayner, K., & Morrison, R. E. (1981). Eye movements and identifying words in parafoveal vision. *Bulletin of the Psychonomic Society*, *17*, 135-138.
- Rayner, K., Reichle, E. D., & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 243-268). Oxford, England: Elsevier.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125-157.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, *39*, 4403-4411.
- Reilly, R. G., & O'Regan, J. K. (1998). Eye movement control during reading: A simulation of some word-targeting strategies. *Vision Research*, *38*, 303-317.
- Russo, J. E. (1978). Adaptation of cognitive processes to the eye movement system. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye Movements and the Higher Psychological Processes* (pp. 89-112). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Salvucci, D. D. (2000). A model of eye movements and visual attention. In *Proceedings of the International Conference on Cognitive Modeling* (pp. 252-259). Veenendaal, The Netherlands: Universal Press.
- Salvucci, D. D., & Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 923-928). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Salvucci, D. D., & Anderson, J. R. (in press). Integrating analogical mapping and general problem solving: The path-mapping theory. *Cognitive Science*.
- Salvucci, D. D., & Anderson, J. R. (in press-b). Automated eye-movement protocol analysis. *Human-Computer Interaction*.
- Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, 26, 1270-1281.
- Suppes, P. (1990). Eye-movement models for arithmetic and reading performance. In E. Kowler (Ed.), *Eye Movements and their Role in Visual and Cognitive Processes* (pp. 455-477). New York: Elsevier Science Publishing.
- Suppes, P., Cohen, M., Laddaga, R., Anliker, J., & Floyd, R. (1983). A procedural theory of eye movements in doing arithmetic. *Journal of Mathematical Psychology*, 27, 341-369.
- Thibadeau, R. H., Just, M. A., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, 6, 157-203.
- Timberlake, G. T., Wyman, D., Skavenski, A. A., & Steinman, R. M. (1972). The oculomotor error signal in the fovea. *Vision Research*, 12, 1059-1064.
- Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Eye Movements and their Role in Visual and Cognitive Processes* (pp. 353-393). New York: Elsevier Science Publishing.
- Volkman, F. C. (1986). Human visual suppression. *Vision Research*, 26, 1401-1416.

Acknowledgments

I am grateful to Erik Reichle, Mike Byrne, Andy Liu, Mike Schoelles, Whitman Richards, John Anderson, Keith Rayner, Sandy Pollatsek, Risto Miikkulainen, and an anonymous reviewer for many helpful comments regarding this work.

Correspondence and requests for reprints should be sent to Dario Salvucci, Cambridge Basic Research, Four Cambridge Center, Cambridge, MA 02142. Email: <dario@cbr.com>.

Footnotes

¹ EMMA and the domain models described here are publicly available at < <http://www.cbr.com/~dario/EMMA> >.

Table 1.

EMMA parameters and values.

Parameter	Description	Setting	Value
K	Encoding constant 1	Estimated	.006
k	Encoding constant 2	Estimated	0.4
T_{prep}	Eye movement preparation time	Estimated	135 ms
T_{exec}	Eye movement execution time	Preset	50+20 ms + 2 ms/deg

Table 2.

Instructed strategies for the equation-solving task.

Name	Strategy
[<i>a B A b</i>]	encode <i>a</i> , encode <i>B</i> , encode <i>A</i> , compute (<i>A/a</i>), encode <i>b</i> , compute (<i>B/b</i>), compute $x=(A/a)(B/b)$
[<i>a A B b</i>]	encode <i>a</i> , encode <i>A</i> , compute (<i>A/a</i>), encode <i>B</i> , encode <i>b</i> , compute (<i>B/b</i>), compute $x=(A/a)(B/b)$
[<i>b A B a</i>]	encode <i>b</i> , encode <i>A</i> , encode <i>B</i> , compute (<i>B/b</i>), encode <i>a</i> , compute (<i>A/a</i>), compute $x=(A/a)(B/b)$
[<i>b B A a</i>]	encode <i>b</i> , encode <i>B</i> , compute (<i>B/b</i>), encode <i>A</i> , encode <i>a</i> , compute (<i>A/a</i>), compute $x=(A/a)(B/b)$

Table 3.

Gaze, first-fixation, and single-fixation durations across frequency classes for data (Reichle et al., 1998) and model.

Class	Mean Frequency (per million words)	Gaze Duration		First-Fixation Duration		Single-Fixation Duration	
		Data	Model	Data	Model	Data	Model
1	3	293	295	248	278	265	288
2	45	272	269	233	256	249	262
3	347	256	256	230	249	243	251
4	4,889	234	220	223	217	235	218
5	40,700	214	215	208	208	216	210
<i>R</i>	–	.98		.96		.96	

Table 4.

Skip, one-fixation, and two-fixation probabilities across frequency classes for data (Reichle et al., 1998) and model.

Class	Mean Frequency (per million words)	Skip Probability		One-Fixation Probability		Two-Fixation Probability	
		Data	Model	Data	Model	Data	Model
1	3	.10	.09	.68	.84	.20	.07
2	45	.13	.19	.70	.77	.16	.04
3	347	.22	.30	.68	.68	.10	.02
4	4,889	.55	.55	.44	.44	.02	.01
5	40,700	.67	.78	.32	.22	.01	.01
<i>R</i>	–	.99		.97		.96	

Figure Captions

Figure 1. Sample screens and eye movements protocols from (a-b) the equation-solving task and (c) the reading task.

Figure 2. Various cases illustrating the control flow of the EMMA model in which (a) encoding and eye movement complete at the same time; (b) a second attention shift cancels the labile stage of the eye movement, resulting in a skipped fixation; (c) a second attention shift does not cancel the non-labile stage of the eye movement; (d) encoding continues after the first eye movement, resulting in a re-fixation of the visual object. The dashed lines represent the separation between the cognitive processor (upper portion) and EMMA (lower portion).

Figure 3. Percentage of trials with four, three, and two gazes in the equation-solving task.

Figure 4. Gaze durations by computation in the equation-solving task. NC gazes involve no computation, 1C gazes involve one computation (i.e., an intermediate value), and 2C gazes involve two computations (i.e., an intermediate value and the final result).

Figure 5. Gaze durations by computation and skipping in the equation-solving task. NC gazes involve no computation; 1C gazes involve one computation (i.e., an intermediate value). NS gazes involve no skipping and occur before a normal gaze on the next equation value; 1S gazes occur before skipping a gaze on the next equation value.

Figure 6. Histograms of deviation ratios in the equation-solving task for (a) the axis parallel to the preceding saccade and (b) the axis perpendicular to the preceding saccade. Deviation ratios are computed as the deviation distance from the actual fixation point to the center of the intended target divided by the distance from the preceding fixation point (i.e., the saccade launch point) to the center of the intended target.

Figure 7. Sample menu from the menu-selection task.

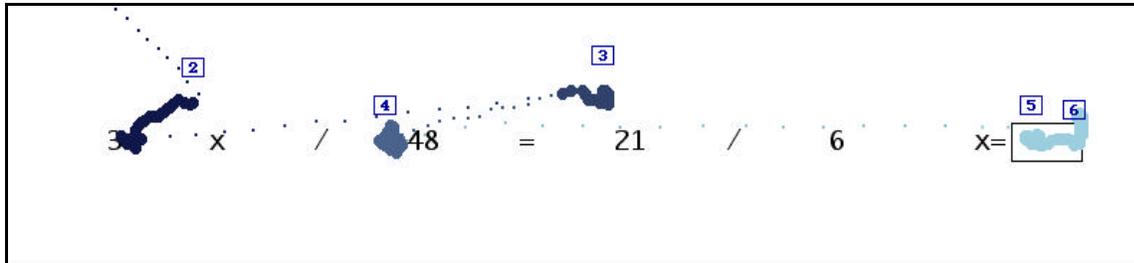
Figure 8. Trial times by target position and menu length in the menu-selection task.

Figure 9. Ratio of first fixations by position and menu length in the menu-selection task.

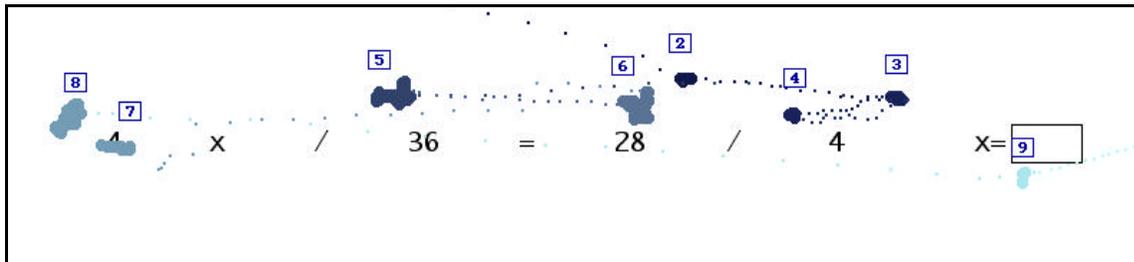
Figure 10. Fixation counts by item position and target position for nine-item menus in the menu-selection task.

Figure 1.

(a)



(b)



(c)

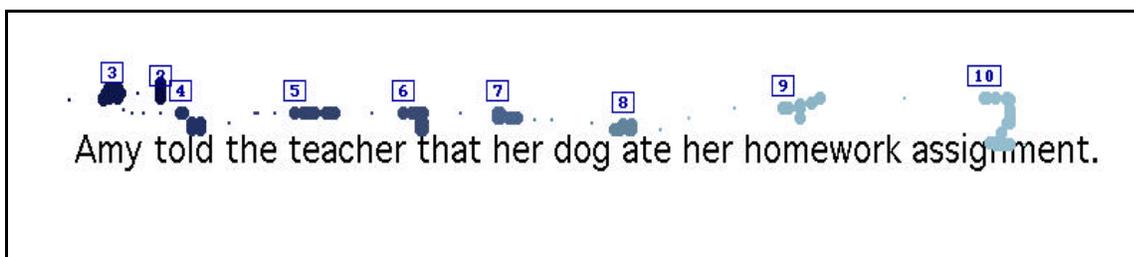


Figure 2.

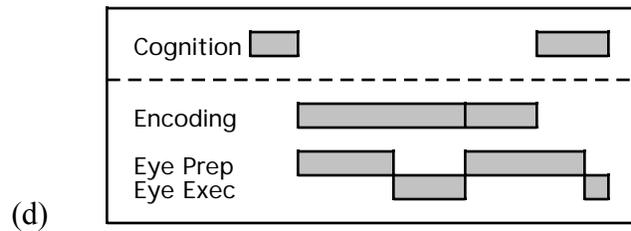
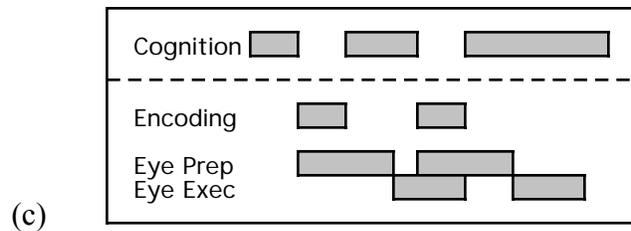
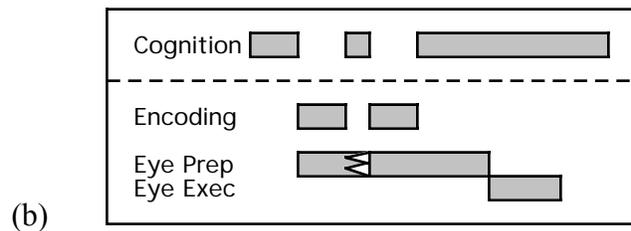
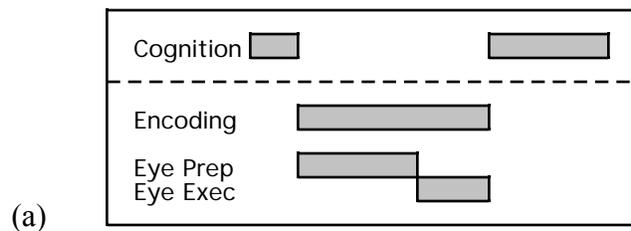


Figure 3.

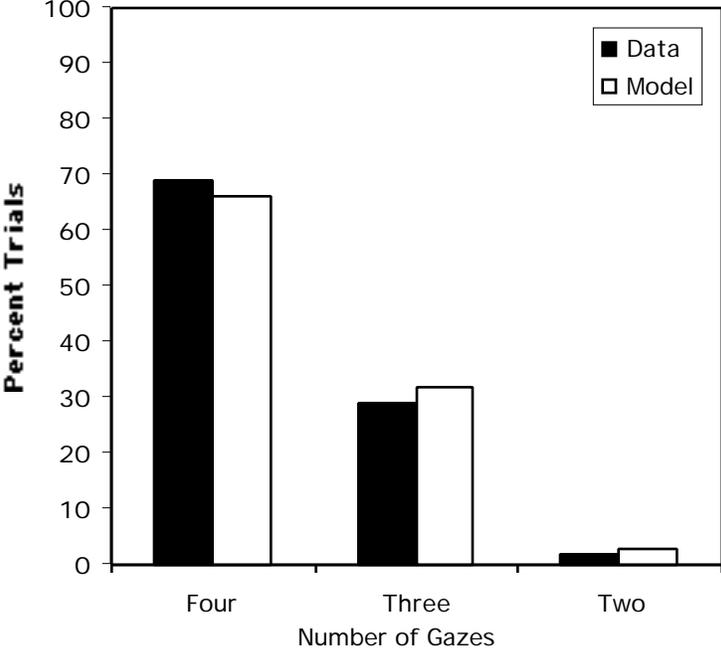


Figure 4.

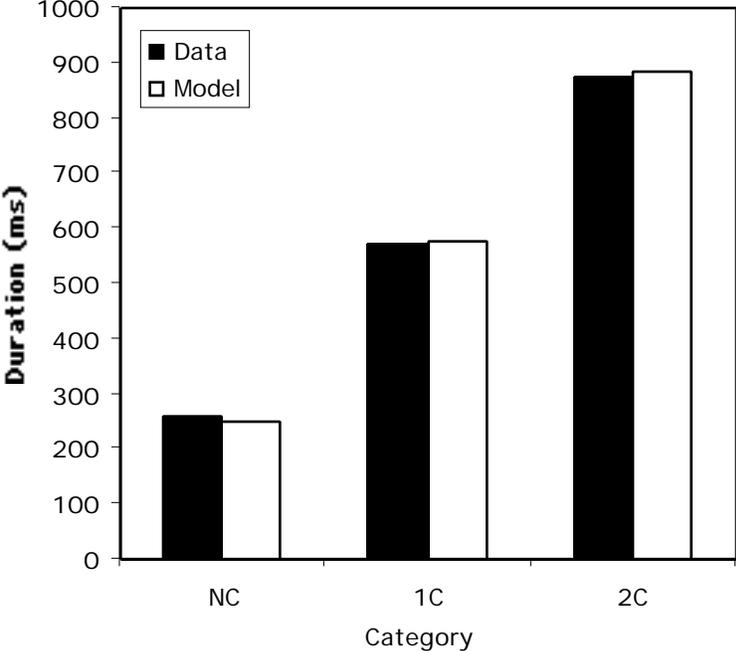


Figure 5.

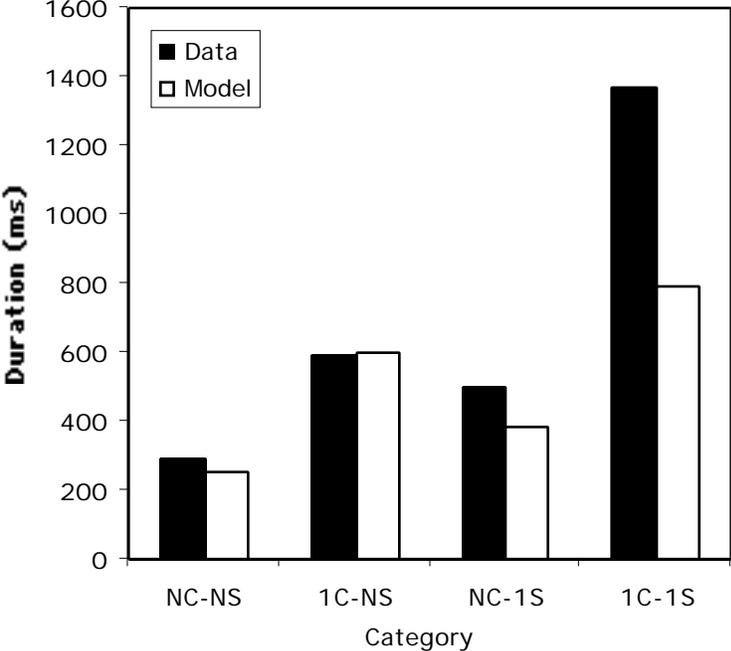
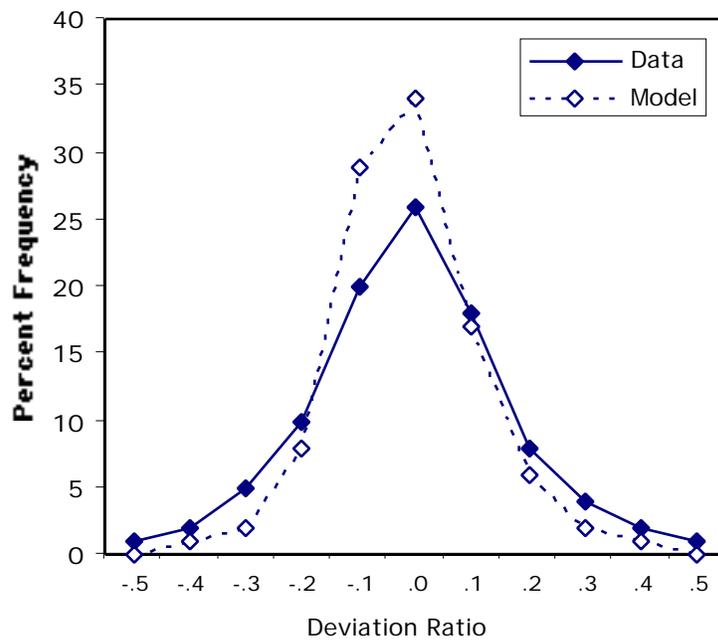


Figure 6.

(a)



(b)

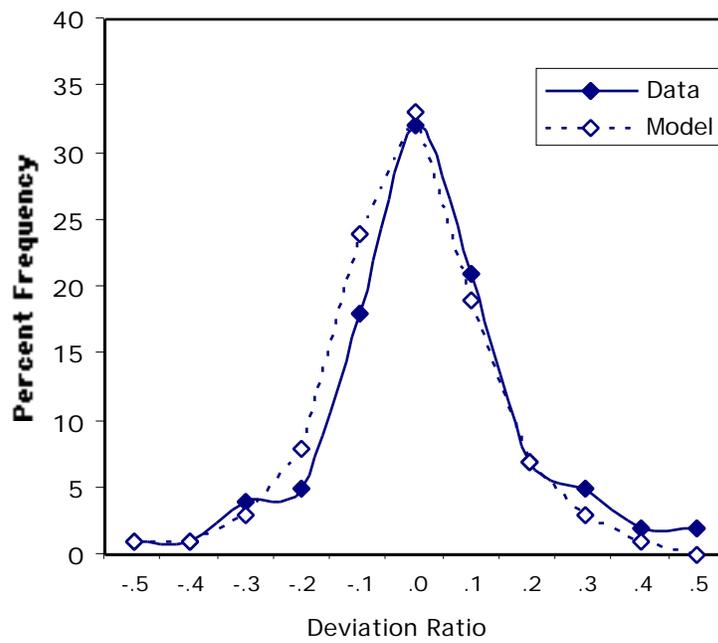


Figure 7.

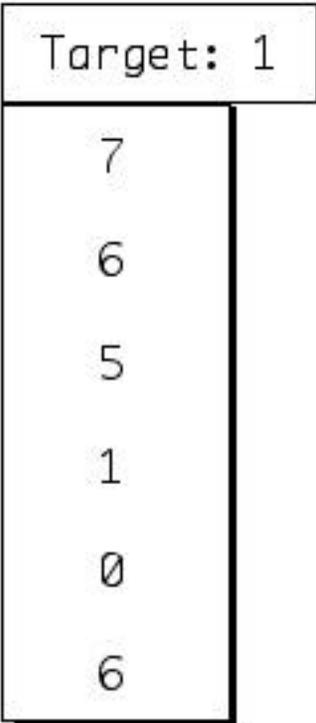


Figure 8.

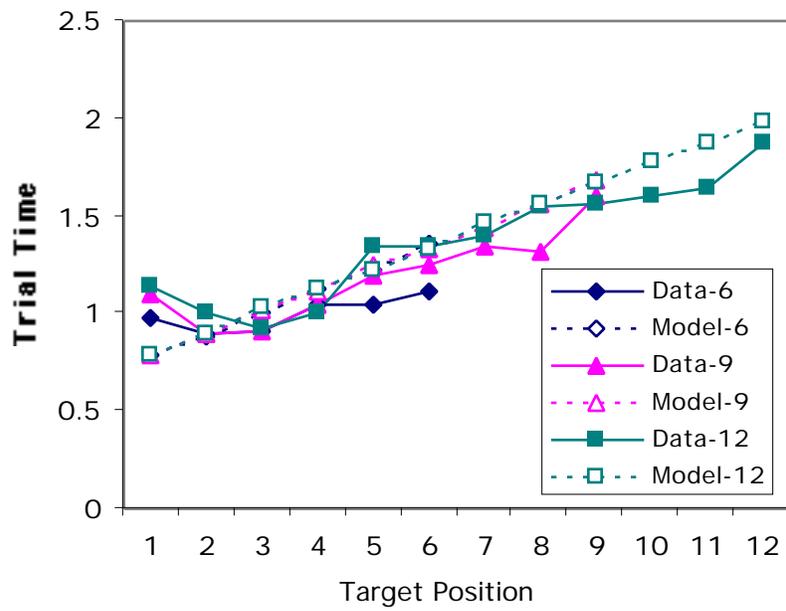


Figure 9.

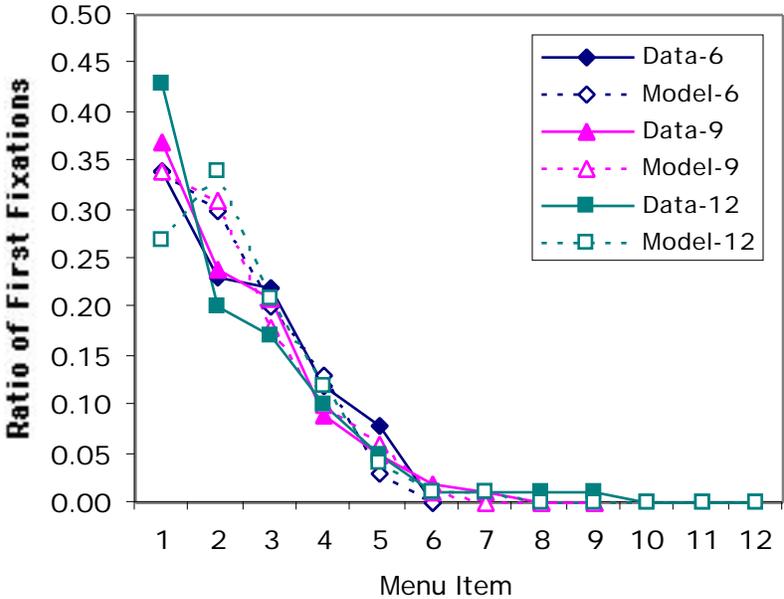


Figure 10.

