# A Model of Eye Movements and Visual Attention

Dario D. Salvucci (dario@cbr.com)
Cambridge Basic Research
Four Cambridge Center
Cambridge, MA 02142 USA

## Abstract

This paper introduces the EMMA model of eye movements and visual attention. EMMA provides a formal model of the temporal and spatial aspects of eye movements as they arise from shifts in visual attention. To implement the model, EMMA is integrated into the ACT-R/PM cognitive architecture (Anderson & Lebiere, 1998) and requires minimal modification of existing ACT-R/PM models. This paper details the EMMA model, theory, and implementation and also demonstrates how the extended framework helps to capture important aspects of behavior in an equation-solving task.

## Introduction

Until recently, cognitive models did not interact with the outside world — they assumed that external stimuli were already encoded in some memory representation and simply acted upon these representations. Today, models that interact with the world through simulated vision, audition, and motor actions are much more common. These models encode stimuli from a simulated environment and produce responses that act upon the environment, sometimes in real time. This ability to interact has given cognitive models an increasing sense of realism in capturing human behavior.

This paper describes the EMMA model of eye movements and visual attention. EMMA (Eye Movements and Movement of Attention) represents an integration of several existing eye-movement models for specific domains into a general model for any problem-solving domain. The model posits that eye movements are initiated by shifts of attention and are sometimes canceled by subsequent shifts. To illustrate and evaluate EMMA, the model is developed within a particular cognitive architecture, ACT-R/PM (Anderson & Lebiere, 1998; Byrne & Anderson, 1998). ACT-R/PM allows for shifts of visual attention as a model encodes components of a visual stimulus, but incorporates a somewhat simplistic model of visual attention and no model of eye movements. EMMA extends ACT-R/PM to these processes and thus enables ACT-R/PM models to account for a richer set of empirical phenomena.[1]

### Eye Movements and Visual Attention

The ability to model eye movements and visual attention independently may seem at first like a subtle and unimportant point. However, modeling visual attention without modeling

[1] EMMA is available on the World Wide Web at < http://www.cbr.com/~dario/EMMA >.

Salvucci, D. D. (2000). A model of eye movements and visual attention. To appear in Proceedings of the International Conference on Cognitive Modeling.

eye movements has a very serious limitation: the difficulty of comparing model predictions to observable data. While we may be most interested in shifts of visual attention, these shifts are hidden from observation; we can only observe the eye movements produced from these shifts. Although eye movements and visual attention are certainly correlated, they do not always correspond directly, especially for complex visual stimuli. This separation of visual attention and eye movements can create great difficulties when attempting to evaluate cognitive models based on their predictions of visual attention.

For instance, consider an arbitrary task that allows for peripheral encoding of visual objects (as do most real-world tasks). In encoding two visual objects, a person can attend to both objects while the eyes remain still at one of the objects or between them. Assuming that the cognitive model must encode both objects, the model would predict some time spent attending to the first object and some additional time attending to the second. However, the eye movements with peripheral encoding would exhibit a single fixation that may or may not be on either object. Thus, even though the model may be a perfectly good model of behavior in the task, the correspondence between predictions and data is poor. These types of problems are exacerbated for tasks with more complex visual stimuli such as reading. While readers must attend to every word in a sentence to comprehend meaning, their eye movements often pass over short and/or high-frequency words (Schilling, Rayner, & Chumbley, 1998). A cognitive model that predicts only visual attention cannot account for these phenomena without some prediction of eye-movement behavior.

### EMMA and Equation Solving

This paper describes EMMA, a model that relates shifts of visual attention to the eye movements they produce. EMMA borrows a number of ideas from existing work to account for various empirical phenomena. The visual attention component of the model addresses how people shift attention to new visual targets and how they process and encode these targets. The eye-movement component of the model addresses how and when people move their eyes to attended targets. Together, these two components produce a rigorous computational theory that allows for closer correspondences between model predictions and observed data. Implemented in the ACT-R/PM architecture, EMMA provides separate ACT-R/PM modules that embody the interactive but distinct separation between visual attention and eye movements.

To demonstrate the theory, we will examine a task in which college undergraduates solve equations of a particular form. Salvucci and Anderson (1998) used this equation-solving task to study how one can interpret eye-movement protocols by means of *tracing* — relating protocols with the sequential predictions of a cognitive model. While the study proved successful at showing the ability of tracing to interpret eye movements, it also showed that students' eye movements often did not correspond to their encoding strategies — for instance, when they encoded equation values peripherally without fixating all values. This paper develops an ACT-R/PM model for the equation-solving task that incorporates EMMA's predictions of eye movements from visual attention and shows how EMMA can help account for a richer set of phenomena than standard ACT-R/PM models.

## The EMMA Model

The EMMA model of visual attention and eye movements borrows and integrates a number of ideas from existing research. EMMA makes extensive use of empirical and modeling

results in reading, particularly Reichle et al.'s (1998) E-Z Reader model. While this model and similar models (e.g., Morrison, 1984; Legge, Klitz, & Tjan, 1997) were designed specifically for reading data, the concepts embodied by the models are applicable to more general cognitive tasks. By being integrating into the existing ACT-R cognitive architecture, EMMA demonstrates how such models and theories can be extended to predict eye movements in other domains.

### Visual Attention

In EMMA, visual attention begins with a command from the cognitive processor to move attention to a given visual object. When the command is issued, EMMA begins the process of encoding the object — that is, recognizing the visual representation and storing it in declarative memory as a more abstract memory "chunk". EMMA posits that encoding time is dependent on two factors: frequency, which measures how often the object is encountered and encoded (Schilling, Rayner, & Chumbley, 1998); and eccentricity, which measures the distance of the object from the fovea (Rayner & Morrison, 1981). EMMA computes the time $T_{enc}$ to encode object $i$ as

$$T_{enc} = c\ \left[-\log f_i\right]\ e^{d_i}$$

where $f_i$ represents frequency, $d_i$ represents eccentricity distance, and $c$ is a scaling constant. The frequencies $f_i$ are defined as values in the range [0,1] that provide some quantification of object frequency — that is, how often the object appears in external world. The eccentricity distance $d_i$ is defined as the distance between the center of the fovea and the object, as measured in degrees of visual angle. Thus, encoding time increases as eccentricity distance increases, but decreases as frequency increases. To add noise to the model, EMMA assumes that encoding time is distributed as a gamma distribution with mean $T_{enc}$ and standard deviation equal to one-third the mean (Reichle et al., 1998).

### Eye Movements

In EMMA, eye movements are initiated by shifts in visual attention. The eye movement is divided into two stages: preparation and execution. When attention is directed to a new target, the model begins preparation of the eye movement. When preparation completes, the model programs and executes the saccade. The separation of the two stages models the fact that people can cancel eye movements soon after an attentional shift, but after some time the eye movement occurs regardless of any changes (Becker & Jürgens, 1979). To set the durations of each stage, EMMA uses duration values estimated for the E-Z Reader model. The preparation stage is assumed to require 150 ms. The execution stage requires 50 ms for motor programming, 20 ms for a saccade, and an additional 2 ms for each degree of visual angle subtended by the saccade (Fuchs, 1971). EMMA also adds noise to preparation and motor programming, again by sampling a gamma distribution with these means and standard deviations equal to one-third the means.

In addition to these temporal characteristics, EMMA models certain aspects of the spatial characteristics of eye movements. Given a saccade to a particular target, the landing point of the saccade is distributed as a Gaussian distribution around the target. The standard deviation of this distribution is 0.1 times the total distance between the original fixation point and the target object, as estimated in previous research (Kowler, 1990). For

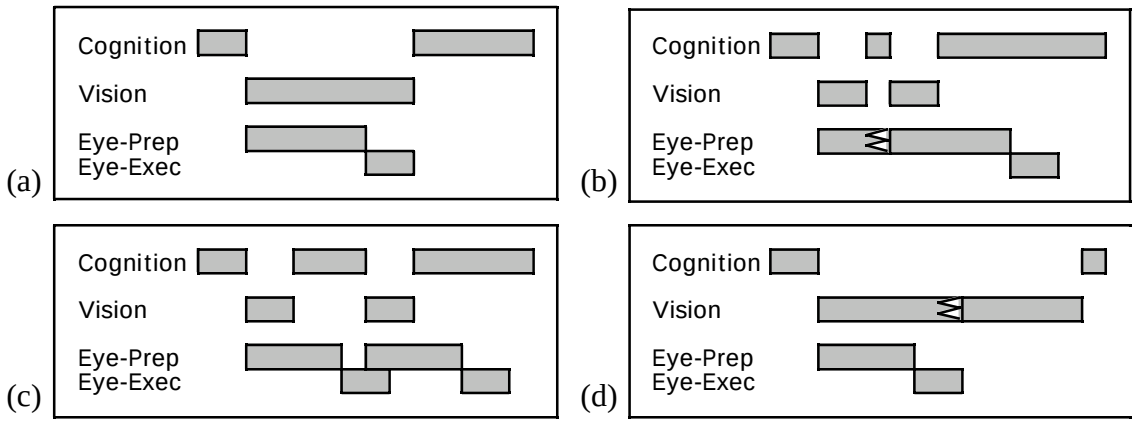


Figure 1: Sample cases of EMMA's control flow.

the sake of simplicity, EMMA does not explicitly incorporate an undershoot bias for saccades, although such a bias could be incorporated into future versions of the model.

**Control Flow**

We can describe EMMA's control flow in terms of four processes: cognition that drives shifts of attention, vision that shifts attention and encodes objects, eye-movement preparation that readies an eye movement, and eye-movement execution that includes both motor programming and execution. These processes run in parallel and, depending on several factors, a number of possibilities can arise. For the purposes of exposition, we now examine the various possibilities and describe the control flow of the model in each case.

In the simplest case, encoding requires the same amount of time as an eye movement. In this case, visual attention works on encoding the object while the eye-movement module runs through each of its two stages. Figure 1(a) illustrates this case, with horizontal bars showing the execution of each module and stage. 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 1(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 1(c). If the eye movement completes before encoding completes, EMMA cancels encoding and restarts it with the new foveal position, as shown in Figure 1(d). This aspect of the model helps to account for behavior when objects are distant from the fovea and encoding time is very long; typically, the restarted encoding process is significantly shorter than the old process due to the decrease in eccentricity.

**Discussion**

Of related existing models, EMMA's formalization of attention and eye movement corresponds most closely to the E-Z Reader model (Reichle et al., 1998). Like EMMA, E-Z Reader incorporates frequency and eccentricity in encoding time and uses a two-stage eye-movement program. However, EMMA and E-Z Reader have at least two important differences. First, E-Z Reader never cancels the encoding process once it is begun; in other words, it has no encoding-time threshold like that of EMMA that limits when encoding

may be terminated and reset. Second, E-Z Reader begins eye-movement preparation to a subsequent target before cognitive processing of the current word is complete, exploiting the fact that the model knows where to fixate next (i.e., the next word). Because EMMA is intended for any domain, it cannot know where to fixate next and requires that the cognitive processor guide attention to the next attended object.

## Equation Solving

We now investigate how the integration of the EMMA model with the ACT-R/PM architecture accounts for eye-movement behavior in the equation-solving domain. This section begins with an overview of the original study (Salvucci & Anderson, 1998) including both reported analyses and additional analyses relevant to the EMMA model. The section then describes a model developed within EMMA and ACT-R/PM that accounts for the observed experimental results.

### Experiment

In the experiment, students solved equations of the form $a\,\mathbf{x} / B = A / b$ by computing $\mathbf{x} = (A/a)(B/b)$. Students completed five total sessions: an initial practice session and four subsequent trial sessions. In each of the four trial sessions, students were instructed to solve the equation using a particular strategy. Each strategy dictated the order in which to encode the values and the order in which to compute the intermediate results. Note that students used only a single strategy during each session to avoid confusion between strategies. Five students participated in the experiment, but one was dropped because of difficulty tracking his eye movements.

The "instructed-strategy" paradigm used in the experiment allows us to know (to a large extent) what cognitive processes were involved in solving the equations. In the original study, this feature of the data was used to test the ability to interpret the observed eye-movement protocols. In the current study, this feature is used to facilitate development of the cognitive model and more rigorously test aspects of the EMMA model. The instructed strategies highly constrain the specification of the ACT-R/PM cognitive model, which predicts the cognitive steps and steps of visual attention performed in task. As such, the paradigm helps us to focus on the actual eye-movement behavior and thus to evaluate rigorously the predictions of the EMMA model.

### Experiment Results

The original experiment discussed aspects of the experiment relevant to strategy use. In this paper we focus on aspects of the experiment relevant to students' eye movements. In particular, we focus on those protocols with four or fewer gazes on the equation elements, comprising 60% of the entire data set. Protocols with more than four gazes presumably involve some amount of review that will not be addressed by our analysis and model.

Figure 2 shows the percentage of protocols with four, three, and two gazes on the equation values. A majority of the protocols (69%) include four gazes — that is, a gaze for each of the equation values. However, there is also a significant number of protocols with less than four gazes — 29% with three gazes and 2% with two gazes. Thus, students sometimes utilized peripheral vision to encode multiple values in a single gaze.

Let us now consider those trials with four gazes on the equation values. The four gazes that comprise each instructed strategy can be classified as follows. First, there is a

start gaze that involves only an encoding and no computation; we label this gaze the S-NC (start-non-computing) gazes. Second, there are two intermediate gazes, one of which involves computation (I-C) and one of which does not (I-NC). Third, there is a final gaze (F-C) that involves computation. For instance, for the strategy [*a B A b*], *a* is the S-NC gaze, *B* is the I-NC gaze, *A* is the I-C gaze, and *b* is the F-C gaze. Similarly, for the strategy [*a A B b*], the gazes in order are the S-NC, I-C, I-NC, and F-C gazes, respectively.

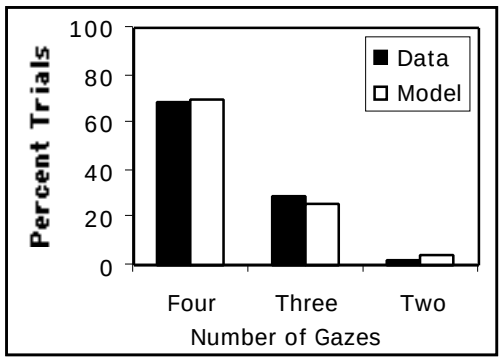


Figure 2: Percentage of trials with four, three, and two gazes on the equation values.

Given this classification, we can analyze the mean gaze duration for each of the gaze positions. Figure 3 shows these mean gaze durations. A repeated-measures ANOVA shows that the effect of position is significant, $F(3,9)=13.75$, $p<.001$. The two gazes that do not involve computation, S-NC and I-NC, have similar durations. The I-C gaze involves one computation (a division of two numbers) and has a mean duration approximately 300 ms greater than the non-computing gaze durations. The F-C gaze involves two computations (one division and the multiplication for the final result) and has a duration approximately 600 ms greater than the non-computing durations. Thus, each computation performed after encoding a value adds roughly 300 ms to the gaze duration on that value.

We can also analyze gaze durations in a way that captures the effects of peripheral encoding. Figure 4 shows the mean gaze durations for four types of gazes classified by two features: whether the gaze involves computation (C) or not (NC), and whether the gaze on the next value in the instructed strategy is skipped (S) or not (NS). This graph includes both three-gaze and four-gaze protocols. The effect of computation, $F(1,3)=109.86$, $p<.01$, skipping, $F(1,3)=32.07$, $p<.05$, and their interaction, $F(1,3)=52.71$, $p<.01$, are all significant. The computation effect arises from the fact that gazes during which computation takes place have longer durations. The skipping effect arises from the fact that when the next gaze is skipped, the encoding occurs during the current gaze, thus increasing the duration of the current gaze. The interaction arises from a rather high value for C-S gazes, caused primarily by gazes that include encoding of I-C and F-C and computation of the intermediate and final results.

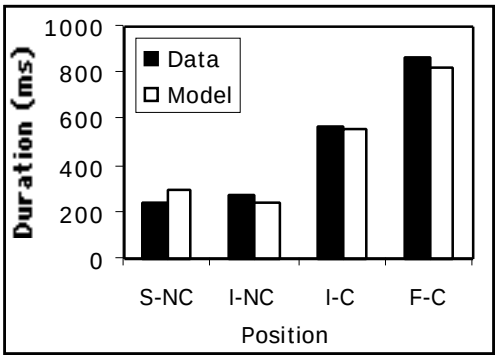


Figure 3: Gaze durations by strategy position for four-gaze protocols.

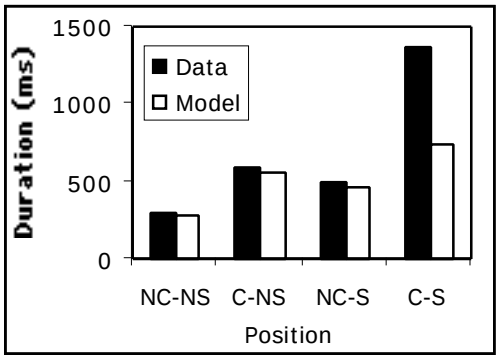


Figure 4: Gaze durations by computation and skipping for three- and four-gaze protocols.

## ACT-R/PM + EMMA Model

The model of the equation-solving task arises directly from the instructed strategies given in the experiment. The model, implemented in ACT-R/PM, embodies these strategies in as simple a way as possible. The standard ACT-R/PM model predicts shifts of attention but not the actual eye movements that accompany them. Thus, this standard ACT-R/PM model cannot predict several results of the experiment, such as skipped fixations and longer durations for fixations before skipped fixations. By running the model with EMMA, we can account for these results with no changes to the basic model.

The model uses two estimated parameters and three preset parameters. EMMA's scaling constant *c* was estimated to have a value of .01, and production strength for all productions was estimated to have a value of 1.2. The effort for all model productions was preset to a value of .01. (This value is smaller than the ACT-R default of .05 to help predict the small durations observed from well-practiced students.) The frequencies of one-digit and two-digit numbers were preset to values of .10 and .01, respectively.

## Model Results

Overall, the equation-solving model running under ACT-R/PM and EMMA provided good fits to most aspects of the observed data. Figure 2 includes the model predictions for the frequencies of different-length protocols, $R>.99$. Because EMMA enables peripheral encoding, the model is able to capture students' use of peripheral encoding to process multiple values in a single gaze.

Figure 3 includes the model predictions for the mean gaze durations by gaze position, $R>.99$. The model computes results just as students do — as soon as possible while looking at the last encoded value — and thus captures the effect of computation on gaze durations: the I-C gaze shows a 300 ms effect for one computation and the F-C gaze shows a 600 ms effect for two computations. These effects arise from the time needed for the model to retrieve the declarative memory chunk that provides the result of the computation.

The model can also predict the effects of skipping on gaze duration. Figure 4 shows the model predictions for three-gaze and four-gaze protocols with respect to computation and skipping, $R=.94$. As in Figure 3, the model exhibits a significant effect of computation. In addition, the model exhibits an effect of skipping the next gaze: when the model encodes two values in a single gaze, thus skipping the next gaze, the duration of this single gaze is greater than it would otherwise be. However, the model does not predict the interaction of computation and skipping present in the large value of the C-S gaze duration; it is possible that while students sometimes both encoded the next value and performed computation with the value during the C-S gaze, the model typically moves its eyes to the next value before computation with the value is complete. This problem may be caused more by the actual model than by EMMA: students seem to purposely maintain gaze on the original target by returning attention to this target, an aspect of behavior not captured by the model.

In addition to the above temporal aspects of behavior, EMMA captures the spatial aspects of student eye movements as well. The data show that the eye movements from an observed fixation to its intended target exhibit Gaussian distributions over the target in both the axis parallel to movement and that perpendicular to movement. The model nicely captures the distributions for these axes, $R=.98$ and $R=.99$ respectively. Unfortunately, due to space constraints, we cannot analyze these data and predictions in detail.

## General Discussion

The incorporation of EMMA into the ACT-R/PM architecture allows all ACT-R/PM models to predict eye movements. EMMA simply extends the existing vision module and adds a separate eye-movement module, but does not change the interface between the ACT-R cognitive model and the ACT-R/PM perceptual-motor modules. Thus, existing ACT-R/PM models that guide visual attention need very few modifications to predict eye movements — these predictions fall directly from the EMMA model of how attentional shifts guide eye movements. We are currently applying EMMA to existing and new ACT-R/PM models to better demonstrate its usefulness and generality.

## Acknowledgments

I am grateful to Erik Reichle, Ken Nakayama, Whitman Richards, and Mike Byrne for insightful comments and helpful suggestions regarding this work.

## References

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

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. 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.

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.

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.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review, 104*, 524-553.

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.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*, 65-81.

Rayner, K., & Morrison, R. E. (1981). Eye movements and identifying words in parafoveal vision. *Bulletin of the Psychonomic Society, 17*, 135-138.

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.

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.

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.