

Scheduling of Eye Movements and Manual Responses in Performing a Sequence of Choice Responses: Empirical Data and Model

Shu-Chieh Wu (scwu@mail.arc.nasa.gov)

San Jose State University and NASA Ames Research Center, Mail Stop 262-4, Moffett Field, CA 94035 USA

Roger W. Remington (R.Remington@psy.uq.edu.au)

School of Psychology, University of Queensland, St Lucia, QLD 4072 AUSTRALIA

Harold Pashler (hpashler@ucsd.edu)

Department of Psychology, University of California, San Diego, La Jolla, CA 92093 USA

Abstract

The seamless and effortless integration between eye movements and cognitive functions signifies tight coordination between eye movement control and the underlying perceptual, cognitive, and motor processes. In this paper we aim to deconstruct the coordination between eye movements and manual responses exhibited in performing a sequence of choice responses through a combination of experimental and modeling methods.

Introduction

Eye movements are integral to human cognition. The seamless and effortless integration between eye movements and cognitive functions signifies tight coordination between eye movement control and the underlying perceptual, cognitive, and motor processes. As the front end of cognitive functions, how the eyes move must bear constraints imposed by the mechanics of the oculomotor system as well as the demand of information processing. But do the eyes fixate and move simply to fulfill the need for information acquisition, or are there other factors involved in determining the scheduling of eye movements? In this paper we aim to shed light on this question by examining the timing of eye movements in performing a sequence of choice responses using a combination of experimental and modeling methods.

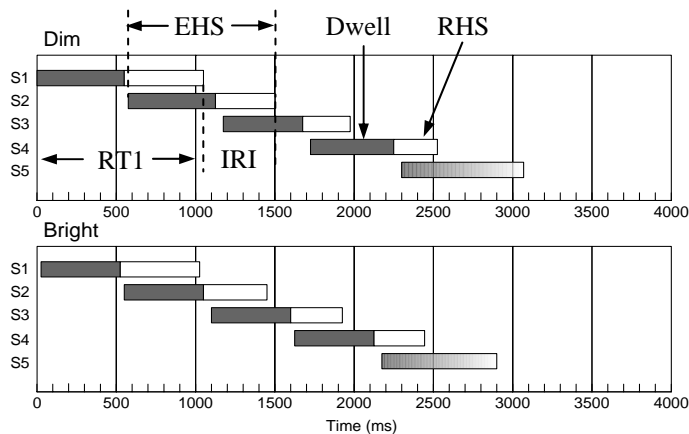
A great majority of research on the coordination between eye movements and cognitive functions focuses on eye movements that occur during reading. In reading, when and where the eyes move are shown to be determined by oculomotor control as well as to a larger extent by visual inputs (i.e., words), in terms of both their physical and linguistic properties (for a review, see Reichle, Rayner, & Pollatsek, 2003). However, most cognitive functions performed in daily life involve not only eye movements but also manual responses or limb movements. Does the need to produce overt manual responses in natural behavior impose further constraints on the scheduling of eye movements?

Efforts to characterize eye movements that occur in natural behavior have focused on activities with well-defined scripts, such as golf putting, driving, tea making, and block-copying (for a review, see Hayhoe & Ballard, 2005). It has been found that in these activities the eyes often move in anticipation of upcoming actions. More fascinating is the tactical timing of anticipatory eye

movements. The eyes appear to move to acquire information just prior to when the information is needed in the action (Johansson et al., 2001). This just-in-time characteristic of eye movement control (cf. Ballard et al., 1995) exemplifies the type of additional constraints imposed by the process of producing manual responses.

As a foray into modeling eye-hand coordination observed in natural behavior, we devised a task complex enough to capture many of the same elements found in natural behavior but simple enough so that its underlying processes can be readily identified (Wu & Remington, 2004). The task was a typing-like task modeled after Pashler (1994). Participants viewed a row of five letters sequentially and responded to each individually. The letters were small and distributed widely so that moving the eyes to fixate each was necessary to performing the task. We evaluated the coordination between eye movements and manual responses through the timing of eye movements, the timing of manual responses, and three derived eye-hand measures: 1) eye-hand span (EHS), which represents the elapsed time between the initial fixation on a particular stimulus to the moment when the corresponding manual response is generated; 2) dwell time, which represents the duration for which fixation is maintained on a particular stimulus; and 3) release-hand span (RHS), which represents the elapsed time between the end of fixation on a particular stimulus to the moment when the manual response is generated. Dwell times and release-hand spans make up eye-hand spans.

Using this simple task we found patterns of anticipatory eye movements commonly seen in natural behavior (Wu & Remington, 2004). Figure 1 shows the pattern of observable events in one experiment that manipulated letter luminance, along with the key dependent measures (Wu & Remington, 2004). The stimuli are listed in the order in which they were responded from top (leftmost) to bottom (rightmost). Horizontal bars reflect the time from fixation to response for each stimulus (S1-S5). RT1 refers to the response time to the first letter (S1). IRI (Inter-Response Interval) is the time between the overt manual responses for each pair of successive stimuli. Fixations are represented by the shaded portion of the bars. Anticipatory eye movements were evident by the fact that the response to a given item was made during fixation on the next item.



Results from this task also revealed emergent properties

Figure 1. The time lines of fixations and manual responses in Dim and Bright luminance conditions (based on the data from Wu & Remington, 2004)

difficult to account for by concatenating processes underlying the constituent discrete responses. Figure 2 plots the results of RT1, IRIs, EHS, and Dwell times measured on each of the five stimuli from the same experiment (Wu & Remington, 2004). As each of the choice response was thought to include identical processes, EHS should remain constant across the series. The empirically observed EHS however decreased across the series. That is, there was a decoupling between eye movement and manual response timing. While the eyes moved across the series at a constant pace, indicated by mostly constant dwell times between S1-S4, manual responses were not produced at a constant rate (with IRI averaged around 450 ms) until after an exceedingly long delay on RT1.

The pattern of regularity in the timing of manual responses (i.e., constant IRIs) was first reported by Pashler (1994), who interpreted it in terms of the central bottleneck

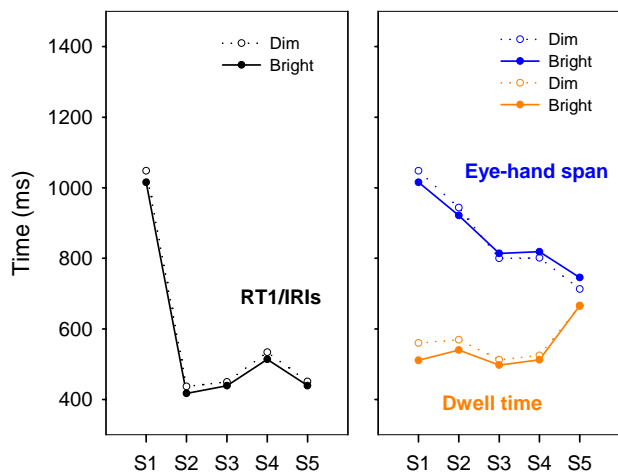


Figure 2. Patterns of RT, EHS, and Dwell results from Wu & Remington (2004)

stage theory. Each choice response is thought to comprise three sequential stages: Stimulus Encoding (SE), Response Selection (RS), and Response Execution (RE). He posited that RS operations on the current item can proceed concurrently with SE operations on subsequent items, and that RE on the current item can proceed in parallel with RS on the subsequent items. RS is the rate limiting operation, and the duration of IRI is a direct measure of the duration of the central RS stage.

The bottleneck theory however could not account for the substantial elevation of RT1. One possible explanation is that RT1 included a cost for performing the sequence that only affects the initial responses and dissipates over time. As the regularity of eye movements appeared to be established from the very beginning, this suggests that eye movements and manual responses could become coupled once the initial preparation cost has completely dissipated. We tested this hypothesis in the present research by examining the effects of sequence length.

Experiment

Experiment investigates the timing of eye movements and manual responses in sequences of differing length. This would allow us to determine if they eventually appear coupled in a longer sequence of responses. It also will allow us to see if the preparation time, reflected in RT1 elevation, is a function of sequence length.

Method

Participants Sixteen undergraduate students recruited from local colleges near NASA-Ames participated in the experiment for course credits.

Apparatus The experiment was carried out on a Pentium 4 PC with a 21-inch monitor. Participants were seated about 28 inches from the monitor. Responses were made using a PC keyboard with fingers of their right hand. Eye movements were monitored using a head-mounted high-speed eye tracker (Applied Sciences Laboratory, Model 501) with eye-head integration function, sampling at 120Hz.

Stimuli and Display The primary stimulus display consisted of a row of nine letters ($0.13^\circ \times 0.26^\circ$) approximately 3.20° apart and centered around the middle of the display. The stimulus letters on each trial were aligned with the leftmost position, with the rest of the positions occupied by small filled squares.

Design and Procedure There were three sequence length conditions (3, 5, and 9). Trials of different sequence length conditions were intermixed. There were a total of 180 trials, 60 in each condition. The trials were administered in 3 blocks of 60. Prior to the experiment participants received 24 practice trials of all sequence length conditions.

Each trial began with the presentation of a fixation cross in the center of the display for 1 second. Then the fixation was erased and a small filled square appeared at the leftmost

stimulus position. Participants were instructed to move their eyes to fixate the small square when it appeared and maintain fixation at that location. The small square remained for 500 ms, followed by a blank interval of 500 ms. Then 3, 5, or 9 letters appeared simultaneously aligned to the leftmost position, with the rest of the positions occupied by small filled squares. Participants were asked to look at the letters one at a time, decide what they were, and make responses accordingly. They were advised to respond as fast as possible but avoid making errors, and to not group responses. The letters were erased after the participant had responded to all of the letters on a trial. The next trial followed immediately.

Results and Discussion

Figure 3 shows the manual response results in the three sequence length conditions. One striking feature was the perfectly aligned results from the three conditions, which all showed the typical pattern of RT1 elevation followed by short and relatively constant IRIs. In addition, in previous experiments IRIs have been found to show moderate increase over the series, with a peak on S4 when the sequence length was 5. It appeared that with extended sequence length IRIs appeared to increase slowly to a new level after every 3 items. It is not clear what caused the pattern of increase.

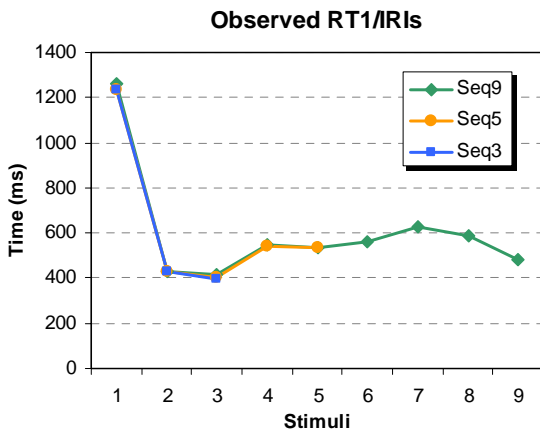


Figure 3. RT1 and IRI results from the present experiment

Figure 4 shows the results of EHS, RHS, and Dwell times in the three conditions. Because results from Sequence 3 and 5 conditions were perfectly aligned with those from Sequence 9 condition, for simplicity only the results from Sequence 9 are plotted. With an extended sequence length, EHS again decreased from the beginning but leveled out after S3, while dwell times stayed relatively constant across the series. RHS also decreased during the first few stimuli and level out after S5. Collectively, the results showed that after three responses, the timing of eye movements and manual responses appeared to be coupled, producing a constant EHS at around 800 ms.

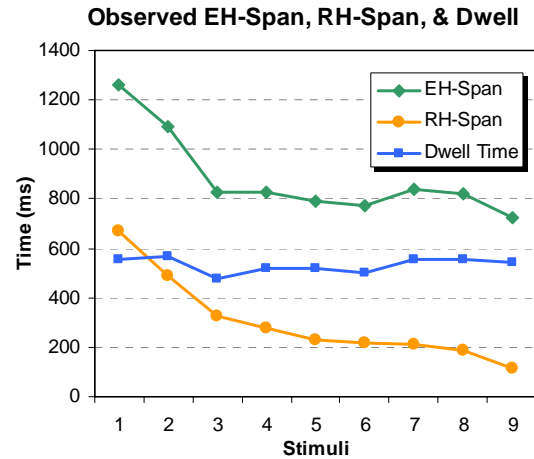


Figure 4. EHS, RHS and Dwell results from the Sequence 9 condition of the present experiment

Model

Previously we reported a model that produced good fits to the data shown in Figure 2 (Remington, Lewis, & Wu, 2006). Here we extend that same model (basic components illustrated in Figure 5) to see how it handles the present data. Here we highlight some basics of the models. A more detailed description can be found in Remington et al., 2006.

Assumptions The model made three key assumptions. First, RS is the rate limiting stage, following the central bottleneck theory. Second, the eyes remain fixated on the current stimulus until SE is complete. Third, the timing of the eye movement is strategically chosen so that SE of the next stimulus is completed at the same time as RS on the current stimulus is completed. We referred to this as the “just-in-time” assumption, since it attempts to minimize wait states in central processing by assuring that perceptual processing is complete as close as possible to when the central processor becomes free.

Model construction The timing of eye movements and manual responses was constructed separately based on their respective hypothesized underlying components. Figure 5 presents the task model of producing a manual choice response. The timing of eye movements was governed by a separate control process that included stages necessary to initiate a saccade (I, denoting Init operator) and to maintain fixation for stimulus encoding (SE) (Figure 6). The

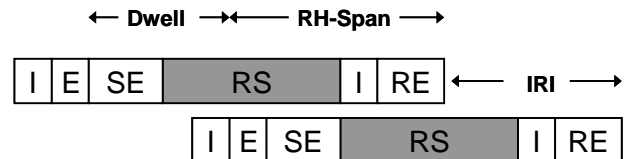


Figure 5. Task model. Processing of each stimulus consists of stimulus encoding (SE, 100 ms), response selection (RS), response execution (RE, 150 ms), and preceded by an eye movement (E, 30 ms). E and RE are preceded by Init operators (I, 50 ms). Dwell, RHS, and IRI are indicated (not to scale)



Figure 6. Task model for processes underlying fixations. SE (150 ms) represented the period of time during which the eyes remained fixated for stimulus processing. Each fixation ended with an Init operator (50 ms) that programs the next saccade.

correspondence between eye movements and manual choice response processing was borne by a Lag parameter that extends fixation beyond what is necessary for stimulus processing to realize just-in-time scheduling. Modeling eye movement timing using a separate control process fulfills the contention that these movements were generated by a lower-level open-loop process not entirely dependent on choice response processes.

Parameters Numerical parameter estimates for several necessary parameters were assigned values consistent with existing literature (e.g., Vera et al., 2005), described in the caption of Figures 5 and 6. The durations for some internal, unobservable states were estimated from data based on theorized processes of the task. For example, RS duration was estimated using averaged IRIs (524 ms). Lag was estimated by first estimating dwell time on a single stimulus, which was estimated by first estimating the total time involved in completing processes on the critical path of the first 8 items, based on the assumption that the eyes remain fixated until the central processor is about to be free. An initial preparation cost was added to RT1, which was derived based the observed RT1 and estimated durations of stages involved in making the choice response.

Simulation Results The model was implemented and Monte Carlo simulations run in the statistical package of R. Model predictions, shown in Figures 7 and 8, were averaged results from 1000 runs. To a large extent, the model again captured the signature pattern of the data. However, the model did not produce the coupling between the timing of eye movements and manual responses found in the observed data between S3-S7, indicated by constant EH-Spans.

General Discussion

The present experiments examined the effects of sequence length on the timing of eye movements and manual responses. Three key findings have direct implications for models of the underlying cognitive mechanisms. First was the striking lack of any sequence length effects on RT1 elevation. This clearly indicates that whatever preparation or start-up costs are reflected in RT1 elevation, they did not accrue from an item-by-item evaluation. That is, if this cost reflects motor planning then the plan is established without consideration of all the items.

Second, sequence length had no effect on asymptotic levels of IRI or dwell time. Again, this suggests that for

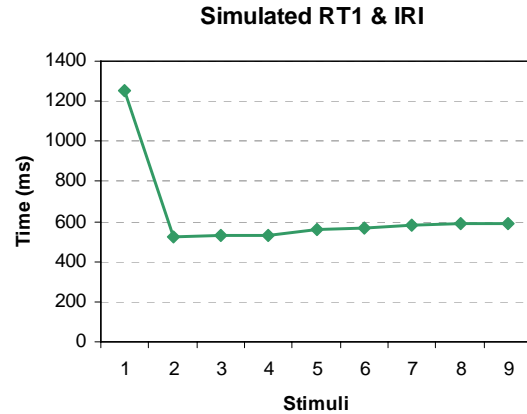


Figure 7. Model predictions for RT1 and IRIs

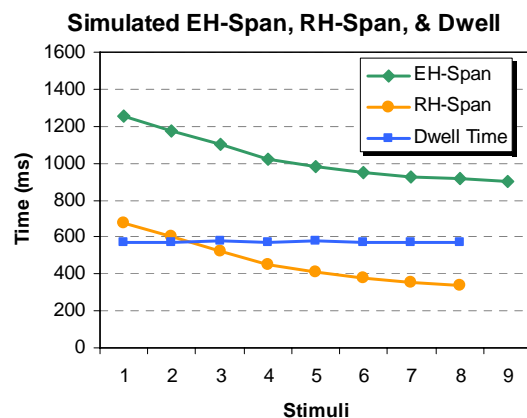


Figure 8. Model predictions for EH-Span, RH-Span, and Dwell time

these regular sequences a simple move-and-respond plan was implemented without examining all items in advance and iterated over items without incurring an item by item cost. It is important to note the simplicity and regularity of our sequences, which may have made these two outcomes possible. It remains to be seen whether the same patterns would be in evidence in visual search with heterogeneous and irregularly dispersed items.

The third outcome of particular importance is the apparent convergence of eye movements and manual responses to a regular tightly coupled phase. This is shown in the EHS curve of Figure 4. The flattening of the EHS after S3 indicates convergence on a more or less constant rhythmic execution of saccades and manual responses. The decoupling of eye and hand responses seen in previous experiments appears to be confined to early portions of the sequence. Indeed, this raises the possibility that the intended strategy in planning to perform the sequence is to achieve this regular rhythmic execution of eye movements and manual responses. The strategy is not reflected in the first few items because of the RT1 elevation, which requires about three items to dissipate. It is this added RT1 processing component that produces the large initial EHS and subsequent decrease that suggested that the eyes and

hands were decoupled. The new empirical evidence instead suggests the possibility that the scheduling of eye movements is constrained not only by visual stimulus processing but also response production. By having participants perform choice responses of various sequence length, we showed that the degree of coordination evolved with the sequence. Coupled eye-hand responses were only found in extended sequences after the initial cost diminishes.

The model that we previously developed to fit the results from 5-item sequences had only moderate success in accounting for the results from 9-item sequences. Specifically, the model continued to predict a decreasing EHS even after 9 items consistent with the decoupled timing of eye movements and manual responses. In addition, the observed data also presented a challenge to the parameter estimates used in the model. Between S3-S7, EH-Spans were about 800 ms when IRIs were about 600 ms. If the duration of IRIs indeed represents the duration of RS, that leaves about 200 ms stimulus encoding, response initiation and execution, which according to model parameters should take 300 ms altogether (100 ms for SE, 50 ms for Init, 150 ms for RE). As it appears that observed IRIs increased with sequence length, it is possible that processes other than RS were involved in determining the timing of manual responses. In future revisions of the model we will explore the possibility that subjects plan the sequence with the goal of maintaining a constant EHS.

The way in which the model accomplishes eye movement scheduling must also be reexamined. Currently, the model estimates the saccade lag parameter by considering the total fixation time on an entire trial. Saccade lag is the parameter that delays the onset of the saccade to attempt to align the end of stimulus encoding on N+1 with the end of response selection on N. Since the total time includes preparatory operations that give rise to RT1 elevation, the estimates of dwell time assume that people consider this when programming sequences and retain these estimates long after the preparatory effects dissipate. In future versions of the model we will explore the consequences of estimating eye movement timing using the average parameter values rather than the total estimated time.

We have described the model as implementing a just-in-time assumption. It is true that the model adjusts saccade lag with the explicit just-in-time goal of having response selection free right when stimulus encoding completes. Because of the RT1 elevation and stochastic stages, just-in-time performance is not achieved in model simulation results. Estimation of the saccade lag directly from average durations of response selection, stimulus encoding, and eye movement latencies may insure better just-in-time performance in practice. Then again, explorations of alternative strategies, such as the eye-hand coupling discussed above, may produce good performance without such an assumption.

To guarantee a just-in-time schedule in model simulation results would require a very different modeling approach.

Our model uses global estimates to set up eye movement timing routines. This contrasts closed-loop control models in which some explicit model construct monitors the momentary progress of cognitive processing and bases the decision to move on the completion of underlying operations (see Reichle, Rayner, & Pollatsek, 2003 for a recent review of modeling approaches to reading). In principle such closed-loop control could achieve more precise timing as eye movement initiation could be adjusted with variations in completion. However, it is unclear whether this would obtain in actual behavior, as the demands of monitoring and deciding on the adjustment could conceivably place increased demands on cognitive processing, interfering with task operations. It may well be that optimal performance is achieved not by precise timing, but by good enough timing done without drawing on central limited capacity resources for either evaluation or execution.

Acknowledgments

This research was supported by grant #FA9550-06-1-0219 from the Air Force Office of Scientific Research, and by grant #DP0666772 from the Australian Research Council.

References

- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representation in natural tasks. *Journal of Cognitive Neuroscience*, 7, 66-80.
- Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 189-194.
- Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, 21, 6917-6932.
- Pashler, H. (1994). Overlapping mental operations in serial performance with preview. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, 47, 161-191.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445-526.
- Remington, R. W., Lewis, R., & Wu, S.-C. (2006). Scheduling mental operations in a multiple-response sequence: Modeling the effects of a strategy to minimize variance in the timing of saccades. In D. Fum, F. del Missier, & A. Stocco (Eds.), *Proceedings of the 7th International Conference on Cognitive Modeling*.
- Vera, A., John, B. E., Remington, R. W., Matessa, M., Freed, M. A. (2005). Automating human-performance modeling at the millisecond level. *Human-Computer Interaction*, 20, 225-265.
- Vickers, J. N. 1992. Gaze control in putting. *Perception*, 21, 117-132.
- Wu, S.-C., & Remington, R. W. (2004). Coordination of component mental operations in a multiple-response task. In S.N. Spencer (Ed.), *Proceedings of the Eye Tracking Research and Applications Symposium 2004*. New York: ACM SIGGRAPH.