Neural plausibility and validation may not be so E-Z

Sara C. Sereno[a], Patrick J. O'Donnell[a], and Anne B. Sereno[b]

[a]*Psychology Department, University of Glasgow, Glasgow, G12 8QB, Scotland, United Kingdom;* [b]*Department of Neurobiology and Anatomy, UT-Houston Medical School, Houston, TX 77030.* **s.sereno@psy.gla.ac.uk p.odonnell@psy.gla.ac.uk anne.b.sereno@uth.tmc.edu**

Abstract: Although the E-Z Reader model accounts well for eye-tracking data, it will be judged by new predictions and consistency with evidence from brain imaging methodologies. The stage architecture proposed for lexical access seems somewhat arbitrary and calculated timings are conservatively slow. There are certain effects in the literature that seem incompatible with the model.

The E-Z Reader model is successful in handling a number of eye-movement phenomena that have been observed during reading. It fares better than other models, although it is not strikingly better than, for example, Glenmore or SWIFT. Notably, there is no comparison across models concerning the functional assumptions about attention shifts or the plausibility of neural mechanisms. It is unlikely that eye-tracking alone and changing parameters of the E-Z Reader model will validate these assumptions. The model has some limitations in its empirical and theoretical justifications (e.g., separation of word identification into two stages, separation of the attentional shift from eye-movement programming), has questionable time frames, and has difficulties with some observations (such as fixation time and word length).

The stage distinction of word identification is crucial to the model because completion of Stage 1 (access to orthographic form) drives the oculomotor system to program the next saccade, while completion of Stage 2 (access to phonological and semantic forms) shifts the spotlight of attention to the next word. A disconnection with independently verifiable temporal estimates for word processing beyond eye-movement measures is a limitation. The danger here is that in setting out to establish a model of eye-movement control, the result may be a model of eye-movement experiments. In the end, a theory of reading will draw on our knowledge of the temporal flow of information across visual and higher-order brain areas, beginning with the first afferent volley that reaches frontal cortex 80 msec post-stimulus and continuing through the top-down feedback loops, which modulate further processing in sensory areas (e.g., Foxe & Simpson 2002). Dense-mapping event-related potential (ERP) and magnetoencephalogram (MEG) recordings provide the requisite spatiotemporal granularity necessary to trace the network of cortical activations millisecond by millisecond. Such research has consistently pushed back estimates of when lexical access, semantic, and top-down contextual effects are first manifest (e.g., Pulvermüller et al. 2001; Rousselet et al. 2002; Sereno et al. 1998; 2003). Assumptions about the temporal course of lexical access should incorporate such finer-grained analyses and use them to test specific predictions that emerge from imputed discrete stages.

One concern about the E-Z Reader model relates to the role of attention. The model decouples saccadic programming from certain shifts of attention. That is, saccadic programming to word$_{n+1}$ begins after the first stage of lexical processing, but an "attention shift" occurs only after a second stage, when the attended word$_n$ has been identified. However, the authors then explain that they are not talking about spatial attention (i.e., attention to spatial orientation), but rather, about attention to "feature," as spatial attention shifts with saccade programming (cf. Sereno 1996). Specifically, they argue that only when an attended word has been identified will attention "shift" to the next word (meaning a shift in some feature space or "analyzer selection"). However, much work in both behavioral studies and physiology has shown that spatial attention improves sensitivity to features at the spatially attended location (e.g., Reynolds et al. 2000; Yeshurun & Carrasco 1998). By artificially separating spatial attention from attention to features, it is not clear or "E-Z" to see how features would be kept separate for proper integration, much less how enhanced stimulus processing from two spatially separate regions is handled in a "strictly serial fashion."

Another concern about the model arises from the implications of its various specifications. For example, Equation 1 can be used to calculate fixation time with regard to word length and fixation location. Given a central fixation point, fixation times for 3-letter and 13-letter words are 95 msec and 115 msec, respectively. It seems implausible that such disparity in word length amounts to only a 20 msec cost in processing. Furthermore, the division of lexical access into two discrete stages seems rather arbitrary: From stage 1 to stage 2, frequency steps down from full to half strength while predictability steps up from half to full strength. Even if one assumes this captures the totality of lexical processing, it would be more parsimonious and neurologically valid to express the decay and growth of these factors as graded functions over time within a single stage. Nonetheless, by summing 90 msec of early visual processing (apparently not accounting for differences in word length) with Stages 1 and 2 (Equations 2 and 3), "word identification time" can be calculated given a word's frequency and predictability. For words that differ substantially in frequency (5 vs. 150 per million) but only slightly in predictability (0.4 vs. 0.6 cloze probability), the resulting times are somewhat vexing: The size of the frequency (34 msec) and predictability (39 msec) effects are comparable, a result that is at odds with the literature in which more robust effects are evident for frequency. In the model, word identification times begin at 148 msec (highest frequency and predictability) and extend to 432 msec (lowest frequency and predictability), a range of 284 msec. However, given that the saccadic programming duration of 240 msec commences at the end of Stage 1 of access and that the longest time to complete Stage 2 is 114 msec, programming duration will always subsume that of Stage 2. Summing Stage 1 and saccadic programming times (across all conditions of frequency and predictability) presumably yields fixation times and these extend from 388 to 558 msec, a range of only 170 msec. Such times seem debatable both in terms of their inflated duration and lack of variability in comparison to fixation times in reading.

Finally, the model seems unable to account for various effects within the literature. Some of these include the interactions of frequency and predictability and of frequency and spelling – sound regularity (e.g., Sereno & Rayner 2000). Also, the "fast priming" eye-movement literature has shown that activation of phonological, semantic, and contextual information occurs very early in lexical processing (Rayner et al. 1995; Sereno 1995; Sereno & Rayner 1992). In a final example, the authors simulate certain lexical ambiguity effects by asserting that the subordinate sense of a (high-frequency) ambiguous word can be treated like a low-frequency word (cf. Sereno et al. 1992). While the word's meaning is of low frequency, its orthographic form is not. And though early access to word meaning may very well drive eye movements, this violates a basic assumption of their model.

We recognize that alternative methods of investigating the time course of semantic processing or the spatiotemporal allocation of attention, such as ERPs or cuing experiments, while offering greater control and finer-grained results, are handicapped with respect to normal reading. However, models of normal reading also make assumptions about lexical access and attention not readily revealed by eye-movement data. The task for the future is to integrate data from complementary, neurally based methodologies.

ACKNOWLEDGMENT

We thank Charles Sereno for his helpful comments.