but to what degree these two factors are involved and how they interact.

Recent modeling work by Engbert and Kliegl (2001) and Reilly and Radach (2003a) can be seen as attempts at reconciling these two views of the reading process. As things stand, however, all current computational models of eye-movement control in reading deal with the process at a relatively shallow level. As pointed out in the target article, one of the real challenges for the next generation of models will be to broaden their coverage to include cognitive and linguistic factors. Unfortunately, as models become more complex, their comparison will become more problematic. The main point of this commentary, therefore, is to make the case for the development of a methodology for the comparison of computational models of eye-movement control in reading.

Our methodological proposals fall under three headings: (1) the facilitation of the comparison of the structural and functional assumptions of competing models; (2) the grounding of models in the neuroscience of vision and language; and (3) the establishment of data sets for model comparison and benchmarking. With regard to the comparison of the structure and function of models, this could be facilitated by using a common implementation framework comprising a set of reusable software components (Schmidt & Fayad 1997). In software engineering terms, a framework is a reusable, "semicomplete" application that can be specialised to produce particular applications or, in this case, particular models. The components would need to be fine-grained enough to accommodate the range of model types and model instances described in the target article. If one could develop an acceptable and widely adopted modeling framework, it would be possible to establish a common basis on which to implement a variety of models. This would make the models more directly comparable not only in terms of their ability to account for data, but also in terms of their underlying theoretical assumptions. The modeling environment could provide a semi-formal language with which a model's structures and processes function could both be unambiguously articulated. This would aid the task of both designing the models and communicating the design to other researchers.

Functionalist computational models, of which E-Z Reader is an excellent example, are inherently underdetermined in terms of their relationship to the brain mechanisms that underlie them. For example, one could envisage a family of E-Z Reader-like models with quite different combinations of parameters and/or parameter values that would be capable of providing an equally good fit to the empirical data (e.g., Engbert & Kliegl 2001). One way to reduce this lack of determinism is to invoke a criterion of biological plausibility when comparing models. We agree with the authors that there is an increasingly rich set of data emerging from the field of cognitive neuroscience which could be used to augment the traditional behavioural sources of constraint on computational models. We believe that models of reading can no longer avoid scrutiny from this perspective. Another, not unrelated, factor in assessing competing models is to take due account of the evolutionary context in which our vision system evolved. Because it evolved for purposes quite different from reading, we need to beware of too-easy recourse to arguments of parsimony, particularly when they are couched solely in terms of the reading process itself. A model with the minimum of modifiable parameters may be parsimonious on its own terms but fail the test of biological realism when compared with, say, a model that comprises an artificial neural network with many hundreds of adjustable parameters. While evolution is parsimonious in the large, when we look at brain subsystems in isolation, such as those involved in reading, we need to be careful how we wield Occam's razor.

Finally, the issue of appropriate data sets with which to test and compare computational models of eye-movement control needs closer attention than has been given to date. The Schilling et al. (1998) data set used to parameterise and test E-Z Reader and several other models discussed in the target article is not particularly extensive. A good case can be made for establishing a range of publicly accessible data sets against which any proposed model can be tested. This would be similar to what has been done, for example, in machine learning, in data mining, and, most notably, in the field of language acquisition (MacWhinney 1995). Furthermore, the corpus of benchmark data should be extended to include a variety of languages, alphabets, and scripts. The more successful models will be those that can readily generalise beyond just one language and one writing system.

Eye-movement control in reading: Models and predictions

Eyal M. Reingold

Department of Psychology, University of Toronto, Toronto, Ontario M5S 3G3, Canada. **reingold@psych.utoronto.ca**
http://psych.utoronto.ca/~reingold/

Abstract: It is argued here that a critical prediction of the E-Z Reader model is that experimental manipulations that disrupt early encoding of visual and orthographic features of the fixated word without affecting subsequent lexical processing should influence the processing difficulty of the fixated word without producing any processing effect on the next word. This prediction is explained and illustrated.

In the target article, Reichle et al. introduce a comprehensive framework for evaluating models of eye-movement control during reading. The authors also provide an updated version of the E-Z Reader model (Reichle et al. 1998; 1999) and argue that the qualitative and quantitative predictions that follow from this model closely match empirical findings concerning a wide range of reading phenomena. Consequently, they contend that the new version of their model, E-Z Reader 7, constitutes the best available computational framework for modeling eye-movement control during reading. The purpose of this commentary is to derive and illustrate a critical and as yet untested prediction that is unique to the E-Z Reader model. The proposed empirical strategy is illustrated in Figure 1 and will be outlined below.

As illustrated in Figure 1, three core aspects of the E-Z Reader model are central to the present proposal: (1) The E-Z Reader model introduces a distinction between two stages of lexical processing: an early lexical processing stage corresponding to the extraction and identification of the orthographic form of the word (L_1), and a late stage involving access to the phonological and semantic forms (L_2); (2) the programming of a saccade to the next word (word_{n+1}) is initiated following the completion of L_1 of word_n; and (3) parafoveal preview of word_{n+1} begins following the completion of L_2 of word_n. Therefore, according to the E-Z Reader model, variation in the duration of L_2 of word_n – $t(L_2)$ – critically determines the duration of parafoveal preview of word_{n+1}.

As shown in Figure 1, the duration of the parafoveal preview of word_{n+1} equals the duration of the interval between the initiation and execution of the saccade to word_{n+1} minus $t(L_2)$ of word_n. In the current implementation of the E-Z Reader model, two variables, word frequency and contextual constraint or predictability, influence the duration L_2 of word_n and consequently should also control the duration of the parafoveal preview of word_{n+1} and the magnitude of any benefit when word_{n+1} is later fixated (e.g., shorter fixations on word_{n+1}, greater probability of skipping word_{n+1}). Consistent with this prediction, greater parafoveal preview benefit on word_{n+1} has been demonstrated when word_n is a high frequency word (Henderson & Ferreira 1990; Kennison & Clifton 1995) and when word_n is highly predictable from the preceding text (Balota et al. 1985). These findings are typically taken to suggest that as the difficulty of foveal processing increases, parafoveal preview benefit decreases.

However, the E-Z Reader model dictates more precise inferences concerning any effects of experimental manipulations of the characteristics of word_n on the subsequent processing of word_{n+1}

https://doi.org/10.1017/S0140525X0346010X Published online by Cambridge University Press

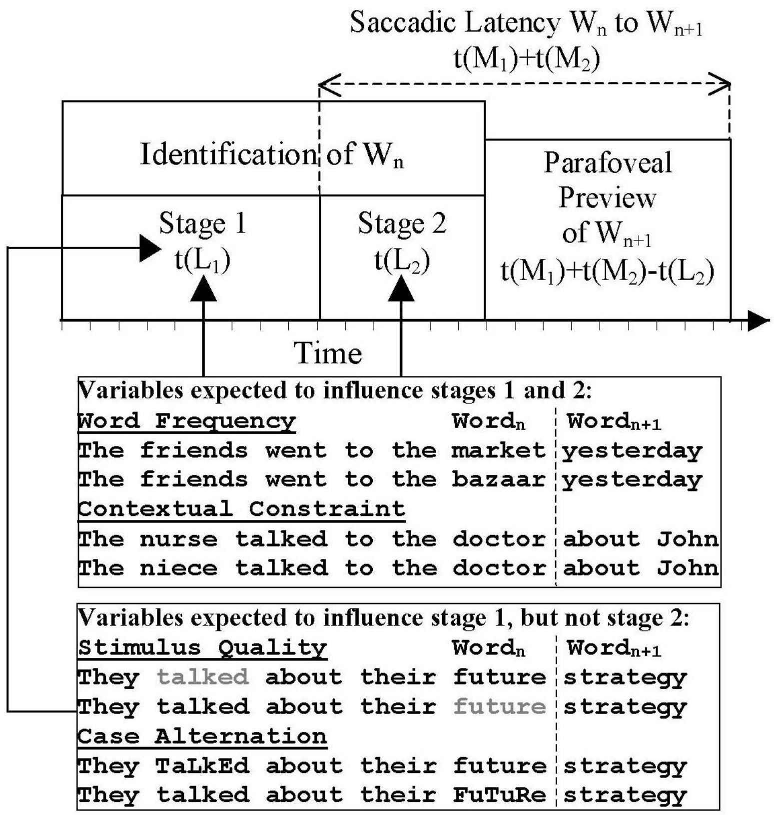

Figure 1 (Reingold). An illustration of the core assumptions of the E-Z Reader model and the methodology proposed for testing this model. The completion of the first stage of lexical processing (L_1) of word$_n$ (W_n) signals the oculomotor system to begin programming a saccade to word$_{n+1}$ (W_{n+1}). On the average, the duration prior to the execution of this saccade (i.e., the saccadic latency) is equivalent to a constant representing the mean duration of two saccadic programming stages ($t[M_1] + t[M_2]$). The duration of the parafoveal preview of word$_{n+1}$ equals the duration of this saccadic latency minus the duration of the second stage of lexical processing $t(L_2)$ of word$_n$ ($t[M_1] + t[M_2]–t[L_2]$). Consequently, experimental manipulations that are expected to influence stage 1 of the lexical processing of word$_n$, but not stage 2 (e.g., stimulus quality, case alternation), are predicted to affect the fixation duration on word$_n$ without influencing the duration of the parafoveal preview of word$_{n+1}$. In contrast, experimental manipulations that are theorized to influence both stages 1 and 2 of the lexical processing of word$_n$ (e.g., word frequency, contextual constraint) have been shown to influence both the fixation duration on word$_n$ and the parafoveal processing of word$_{n+1}$.

(i.e., parafoveal preview effects or spillover effects). Specifically, such effects on word$_{n+1}$ are predicted *if and only if* an experimental variable influences L_2 of word$_n$. This has two vital implications for the empirical validation of the E-Z Reader model. First, variables other than frequency and predictability, such as syntactic difficulty, that have been shown to modulate the magnitude of the parafoveal preview benefit (Henderson & Ferreira 1990) can be inferred to influence the duration of L_2, suggesting possible extensions to the E-Z Reader model. Second and more important, variables influencing L_1, but not L_2 of word$_n$, while modulating the difficulty of the lexical processing of word$_n$, should not affect the magnitude of any processing benefit when word$_{n+1}$ is later fixated. This marks a clear departure from the hypothesis that a substantial increase in the difficulty of foveal processing invariably results in a decrease in the parafoveal preview benefit (Henderson & Ferreira 1990).

I argue here that searching for variables that influence the processing difficulty of word$_n$ without producing any processing effect on word$_{n+1}$ is a critical test of the E-Z Reader framework. A closer examination of the L_1 versus L_2 distinction proposed by E-Z Reader 7 suggests several potentially promising experimental manipulations. Essentially, the L_1 versus L_2 distinction assumes an early lexical processing stage corresponding to the extraction and identification of the orthographic form of the word and a late stage involving access to the phonological and semantic forms. Accordingly, a disruption early in the word recognition system when visual features are encoded and abstract letter identities are computed should be expected to influence L_1, but not L_2. Two manipulations that are generally believed to disrupt early encoding of visual features are illustrated in Figure 1: *stimulus quality* (for a review see Borowsky & Besner 1993) and *case alternation* (for a review see Mayall et al. 1997). However, it has been recently suggested that unlike the former, the latter variable may also influence post-encoding lexical processing (Herdman et al. 1999) or attentional processing (Mayall et al. 2001). It is important to note that the description of the manipulations of stimulus quality and case alternation is meant to merely illustrate, rather than exhaustively detail, the potential value of the proposed research strategy for the of study eye-movement control during reading in general, and the empirical validation of the E-Z Reader model in particular.

ACKNOWLEDGMENTS

This research was supported by a grant to Eyal Reingold from the Natural Science and Engineering Research Council of Canada (NSERC). The author is grateful to Keith Rayner, Elizabeth Bosman, Ava Elahipanah, and Colleen Ray for their helpful suggestions and input. Correspondence concerning this commentary should be sent to Eyal M. Reingold, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3. E-mail: reingold@psych.utoronto.ca.