Vision Research 49 (2009) 1916–1918

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Reply

Non-canonical reading: Reply to Rayner, Pollatsek, Liversedge and Reichle (2008)

1. Introduction

Rayner, Pollatsek, Liversedge and Reichle offer a spirited critique of our paper (Kennedy & Pynte, 2008). Their argument is that our data were obtained selectively, that the effects reported are not typical of normal reading and, in any case, do not pose particularly severe theoretical problems for serial models of reading like *E–Z Reader*. In this rebuttal we follow the organisation of their Comment and deal with three issues: (1) the nature of the phenomena; (2) the degree to which these represent a genuine challenge to current models of eye movement control; and (3) the relationship between empirical data and computational models in this domain.

1.1. What are the phenomena?

With regard to our data on adjective-noun order in French, Rayner et al. make two points. First they argue that the cases reported represent only a small percentage of the data set as a whole. It is not explained what follows from this. We claim nothing follows from it. The data reported represent *all* the cases present in a sizeable corpus (i.e. 263 cases). Self-evidently, the phenomenon is relatively rare. But this fact would not, for example, preclude carrying out experiments on presentation order over word pairs that are infrequent in a particular language (e.g. Lukatela, Kostic, Todorovic, Carello, & Turvey, 1987). Indeed, experiments are often carried out on linguistic constructions that are extremely rare in the written language, such as sentences containing reduced complements without punctuation (see Mitchell, Xingjia, Green, & Hodgson, 2008, for a discussion). It is certainly legitimate to ask whether our results generalise across the set of items used (in this case *all* the relevant items); we do this by reporting by-Participants and by-Items analyses. A second criticism is that fixation duration associated with reverse-order inspection is relatively short. We accept this is the case, and we offer an account in the paper for why this absolute difference is found. However, the point remains that there is a complete absence of any interaction between Order and Syntax. That is, regardless of the absolute duration, it seems to make no difference at all whether an out-of-order sequence is legal or not in the language.

Rayner et al. raise similar points about sample size considering our data on canonical fixation order in the Dundee Corpus as a whole. We examined all strings greater than eleven fixations long. We readily accept that including all strings susceptible to our technique of analysis (>3 fixations) would increase the proportion of apparent canonical reading. But this is because virtually any patterns of fixation over short sequences of words is licensed by the serial model. Nonetheless, we strongly dispute the claim that our data rest on "...a selective sample of long and possibly complex sentences in which the syntax and meaning may not be clear." Taking the data set as a whole, we report more than twenty thousand cases: it is unreasonable to dismiss this as excessively selective. Average string length was around 20 words with an average standard deviation around eight, fairly typical of newspaper text. In summary, we resist the implication that we are dealing with fringe effects or artifacts arising from encounters with poorly constructed sentences. In fact, the force of the criticism with regard to selection escapes us. A credible and robust model of reading should not be particularly embarrassed by parameters of a large and representative data set like those summarised in our Table 1.

Rayner et al. return to the issue of selectivity in a discussion of the data presented by Hogabaom (1983). We agree these data represent an important early examination of the issue addressed in our paper, namely, the consequences of non-canonical fixation order. Unfortunately, they also demonstrate that it is impossible to handle the issue by means of a system of classification, however ingenious. Hogaboam examined the data of 30 participants reading a passage only 417 words long (our data sets were, of course, very much larger). Even for this short text, and using a taxonomy of 15 different patterns, he was only able to capture slightly less than 60 percent of all possible patterns. Rayner et al. observe "It is not clear what happened to the other forty percent of the sequences." We suggest they represent patterns that defeated the taxonomy adopted. It would be idle to present a rival set of speculative estimates from these data, because, broadly speaking, we agree with Rayner et al. on this point. But, with respect to our critics, they appear to underestimate the consequences of their own argument. Viewed as a pattern, each systematic deviation from the canonical may indeed be relatively rare, but the cumulative effect of many such deviations can be great. Indeed, this is the essence of Hogaboam's own argument[1]:

"*Eye movements during skilled reading have been characterised...as largely consisting of fixations on almost every word followed by saccades that take the eye forward to the next word. Although it is true that this pattern was the single most prevalent pattern, it occurred only 22.7% of the time! Models of reading assuming this characterisation might be disregarding over three-fourths of the normal eye movement data...The point to be taken from this is that it is inaccurate to characterize skilled reading as a process of moving one's eyes forward from one word to the next with occasional regressions*". (Hogabaom,1983, pp. 314–315).

We acknowledge that the serial models developed over the intervening 25 years have adopted more and more complex assumptions to account for the "missing three-fourths" of the data. In the case of the *E–Z Reader* model, with great success. Nonetheless, it could well be that a serial mechanism is simply not the best way of characterising the reader's behaviour.

[1] On the basis of the relative frequency of patterns defined as non-canonical on our criteria, Hogaboam's data suggest that strictly canonical reading of sequences longer than 20 words would *never* occur.

doi:10.1016/j.visres.2008.12.013

1.2. Problems for a serial model of eye movement control?

In the next section of their critique Rayner et al. set out to examine whether our data "present a challenge for current models of eye movement control in reading". In fact, their arguments do not deal with "current models" in the plural, but are restricted to a discussion of whether the (serial) *E–Z Reader* model, as currently implemented, can account for our data. In particular, they do not discuss the problems the data present to parallel models such as SWIFT (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005). We were initially puzzled by this, because our paper plainly argues that non-canonical reading poses a challenge to *both* serial and parallel models. The reason for dealing with only half of the theoretical challenge is then made explicit. Kennedy and Pynte, we are told, "are trying to address... whether encoding of words in text is serial (presumably usually left-to-right in English or French texts) or a parallel encoding of two or more words at one time." It is true we have elsewhere claimed that parafoveal-on-foveal effects point to the parallel processing of adjacent words (Kennedy & Pynte, 2005). But the notion of parallel encoding is not the focus of the paper under discussion here. Its purpose is set out in its opening sentence; namely, the consequences of inspecting words in text in a temporal order that violates their spatially-defined word order. Our central thesis is that it is the spatial disposition of text that gives it its necessary order. In a serial model, the only way of preserving this essentially spatial order is to examine the words (overtly or covertly) in a temporal sequence that honours spatial succession. We make no claims with regard to "encoding[2]". On the contrary, it is the serial model that relies on an explicit, but ambiguous, link between encoding order (we take this to mean the underlying temporal order) and spatial succession. One example of the ambiguity is the assertion in the quotation noted above that serial encoding is "presumably usually left-to-right". Quite obviously, the terms "left" and "right" refers to a spatial relationship: to describe this as "serial" is to smuggle the property of spatial succession into the model. The authors of the *E–Z Reader* model are aware of this. Footnote 13 (page 40) in Pollatsek, Reichle, and Rayner (2006) states:

We realized when we were writing this section that there is a figure that we have used in several expositions of our model (e.g. Reichle, Rayner, & Pollatsek, 2003, *Fig. 3), that shows an "attentional spotlight" that is basically the area that is identified as being the field of useful vision from moving window experiments. This metaphor is potentially misleading because it suggests: (a) a separate attentional mechanism than the one we are using in our simulations, and (b) that this whole area is attended simultaneously. Instead, this area is meant to represent information that may be attended to at some time during a fixation; that is, for words in this region, there is a probability greater than zero that (the 1-word) attentional spotlight will be directed on them on some fixations.*

It is fundamental to the rationale of a model like *E–Z Reader* that a "whole area" cannot be attended simultaneously. It cannot even be represented simultaneously, because to do so would be tantamount to providing parallel spatial tags inside the beam of the attentional spotlight.

In summary, regardless of whether "encoding" is serial or parallel, order violation takes on a special significance for serial models because adverse consequences are predicted if two adjacent words are encoded "other than going left-to-right" (Pollatsek et al., 2006). In the SWIFT model the concept of a correct temporal sequence is meaningless because target words may be inspected in any order. It is the case that our data (and those of Hogaboam) suggest that irregular inspection is very common. But this observation does not, in itself, lead to a claim that words can be (or are) "encoded in parallel": Indeed, we explicitly endorse the claim made by Pollatsek et al., that the "*run home*, *home run*" question poses a problem for the SWIFT model (although the authors of SWIFT may well contend that the issue lies outside the boundary conditions of their model).

The part of the comment dealing with the nature of the challenge presented by our data, misrepresents our principal claim. We claim non-canonical reading order presents a challenge to both serial and parallel models. We argue both are incomplete because they do not explicitly allow for the encoding of spatial information. Nonetheless, we wish to make a crucial distinction. In the case of a parallel model like SWIFT, incorporating some kind of spatial tagging appears relatively tractable. That is, the concept of spatial tagging is not self-evidently alien in a model proposing a spatially distributed field of activation undergoing a continual process of dynamic change. The authors of SWIFT have speculated on a possible solution to the problem and, as noted in our paper, the model already predicts the occurrence of large regressive saccades to incompletely processed words. In contrast, apart from the role played by acuity in early visual processing, incorporating spatial tagging (an essentially parallel concept) into a model like *E–Z Reader*, would render its essentially serial character otiose. We believe this is why, in another context, empirical data that do, in fact, appear to support "parallel encoding" must be dismissed as *necessarily* artifactual (e.g. a by-product of selection bias, task bias, measurement error, "zoning out", mislocation, etc.). The difficulty with this approach is that non-canonical reading order is quite common yet appears to have minimal consequences for comprehension. It is this fact, not particular claims with regard to parallel encoding, that constitutes the challenge.

Rayner et al. next consider the degree to which regressions to previously unfixated words represent a challenge to a strictly serial model like *E–Z Reader*. It is common ground that readers misconstrue text and may need to re-analyse it (indeed, many psycholinguistic experiments are designed to produce exactly this outcome). Rayner et al. argue that readers may use regressions to correct faulty initial encoding of a word. They go on to comment that regressions "targeted to that word" may themselves be subject to error. But this criticism is beside the point. We obviously accept that mis-processing a word or mis-parsing a sentence leads to comprehension difficulty and that some means must exist to carry out a repair. We also agree that these means may be irrelevant to the question of whether the encoding of words is sequential or in parallel. Where we part company with our critics is in the use of the word "targeted". Neither serial nor parallel models offer an account of *how* this target is computed. In particular, there is no mechanism in a strictly serial model that allows spatial information to be retained. How, or whether, words can be encoded in parallel is an interesting question and we have some sympathy with the argument expressed in Pollatsek et al. that the concept of parallel encoding is seriously under-specified. But the claim, "of course, they may go back and glance at the passage they skipped if they subsequently decide they were wrong about this." is even more problematic. Such re-inspecting saccades can only be deployed if the reader has retained the necessary spatial information to execute them. The point at issue is not whether regressions are due to mis-targeting, comprehension failures, or failed guesses (we accept they probably are); it is *how* these targets are computed in the first place. The *E–Z Reader* model is silent on this issue, not because additional tweaks to the model can accommodate the data, but because a strictly serial model cannot, in principle, code spatial position.

In this section of their critique, Rayner et al. also take us to task for the claim that non-canonical fixation sequences have little

[2] The word "encoding" appears 15 times in the Comment provided by Rayner et al. and zero times in our paper.

effect on comprehension. We concede that our argument on this point is indirect: we have no "on-line" index of comprehension, although we do present "litmus tests" to deal with this perceived weakness. There is a considerable literature suggesting that wrap-up reflects processing difficulty, and the pattern of effects we report relating to relative clauses and commas is consistent with this. We can only reiterate that if non-canonical reading order leads to major problems of comprehension, there is no evidence to support the proposition in our data.

1.3. Models and data

In this section, Rayner et al. discuss the relationship between computational models and empirical data. We agree that it is important to determine the boundary conditions of any model. Empirical data that address issues outside these boundary conditions should not be presented as a "challenge". Nonetheless, the argument that most of the issues surrounding non-canonical reading are outside the legitimate scope of *E–Z Reader* is too strong. Disruptions may be due to high-level processing (e.g. computing syntactic relations), a concept currently outside the scope of the model. But high-level processing operations are, we presume, built over lower-level "encoding" operations and these are patently within the model's scope. Do Rayner et al. wish to claim that identifying words in sequential order is not a prerequisite for syntactic and semantic processing? We thought that the need for lexical processing to be performed in sequential order was the knock-down argument against parallel models (what we refer to above as the "*run home*, *home run*" question). We argue that strict sequential order is not necessary for comprehension (Rayner et al. appear to concede this point, although we may quibble about the definition of "comprehension"). The *E–Z Reader* model denies itself the mechanism that will permit this to occur: spatial coding. It must do this, because if coded spatial location is allocated to each word (or to some words), sequential order is no longer necessary and one important justification for a "strictly serial" model is removed (of course, visual acuity plays a role in a serial model but its role is defined as a perceptual constraint not as something determining spatial memory.)

We will restrict ourselves to a very brief comment on the broader questions raised with regard to modelling, partly because they go well beyond the present debate and partly because the issues have been rehearsed in detail elsewhere (Jacobs, 2000). Two points can be made. First, in general, current models of eye movement control in reading are far more sophisticated than the techniques available for distinguishing between them. In particular, tests aimed at discriminating between different classes of model are often surprisingly weak, and sometimes offer little more than an invitation to see apparent similarities in the form of distributions. Second, if the data-fitting purpose of a particular model is satisfied by the proliferation of parameters, a situation may arise (or may already have arisen) where models with radically different psychological assumptions account for all the available data (Radach & Kennedy, 2004). Perhaps we can find common ground in the modest claim that models should be capable of generating testable predictions and should be capable of being falsified. The *E–Z Reader* model and the SWIFT model score well on both fronts, because their assumptions are very explicit and their states are "psychologically committed". Where we clearly disagree with our critics is the extent to which the present data do, in fact, falsify crucial predictions.

Acknowledgments

Alan Kennedy is supported by Grants from the UK Economic and Social Research Council (Grant Number RES 000–22–2585).

References

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*, 621–636.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813.

Hogabaom, T. W. (1983). Reading patterns in eye movement data. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 309–356). New York: Academic Press.

Jacobs, A. M. (2000). Five questions about cognitive models and some answers from three models of reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 721–732). Oxford: Elsevier.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45*, 153–168.

Kennedy, A., & Pynte, J. (2008). The consequences of violations to reading order: An eye movement analysis. *Vision Research, 48*, 2309–2320.

Lukatela, G., Kostic, A., Todorovic, D., Carello, C., & Turvey, M. T. (1987). Type and number of violations and their grammatical congruency effect in lexical decision. *Psychological Research, 49*, 37–43.

Mitchell, D. C., Xingjia, Shen, Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the selective reanalysis hypothesis. *Journal of Memory and Language, 59*, 266–293.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the *E–Z Reader* model: Exploring the interface between cognition and eye movement control. *Cognitive Psychology, 52*, 1–56.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues and an agenda for future research. *European Journal of Cognitive Psychology, 16*, 3–26.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E–Z Reader model of eye movement control in reading: Comparison to other models. *Brain and Behavioral Sciences, 26*, 445–476.

Alan Kennedy
School of Psychology, The University of Dundee, Dundee, Tayside DD1 4HN, United Kingdom
Laboratoire de Psychologie et Neurosciences Cognitives, Boulogne-Billancourt, France
E-mail address: a.kennedy@dundee.ac.uk

Joël Pynte
Laboratoire de Psychologie et Neurosciences Cognitives, Boulogne-Billancourt, France