This is a preprint of an article whose final and definitive form was published in

Journal of Experimental Psychology: General 2007, Vol. 136, No. 3, 530–537

DOI: 10.1037/0096-3445.136.3.530

Toward a Perceptual-Span Theory of Distributed Processing in Reading: A Reply to Rayner, Pollatsek, Drieghe, Slattery, & Reichle (2007)

Reinhold Kliegl

University of Potsdam, Potsdam, Germany

Running Head: Distributed processing in reading

Address for correspondence:

Reinhold Kliegl (email: kliegl@rz.uni-potsdam.de)

Department of Psychology, University of Potsdam

Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

phone: +49 331 977 2868; fax: +49 331 977 2793

Abstract

Rayner et al. (2007) argued that our corpus-analytic evidence for distributed processing during reading should not be accepted, because (1) there might be problems of multicollinearity, (2) the distinction between content and function words and the skipping status of neighboring words was ignored, and (3) there are inconsistencies with experimental results. Re-analyses with linear-mixed effect models demonstrate that (1) regression coefficients are stable across nine samples, (2) lexical status and skipping status (and their interactions) are highly significant but do not account for the effects of word frequency for content and for function words, and (3) there is strong evidence for lexical processing of content words while fixating function words to the left of them. A critical result about fixation durations prior to skipped words is replicated in an experiment. The distinction between "correlational analyses" and "well-controlled" experiments and questions about generalizability of results are discussed. I argue for a complementary role of corpus analysis, computational modeling, and experiments in reading research.

Rayner and colleagues have contributed much knowledge about eye movements in reading with experiments (see Rayner, 1998, for a review) and with the E-Z Reader model, which they evaluated with corpus data from reading of sentences (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998; Schilling, Rayner, & Chumbley, 1998).[1] In their Comment on Kliegl, Nuthmann, and Engbert (2006; hereafter Kliegl et al.), Rayner, Pollatsek, Drieghe, Slattery, and Reichle (2007; hereafter Rayner et al.) suspected that our evidence for distributed processing during reading fixations is due to (1) multicollinearity between predictors, (2) the inclusion of fixations on function words, or (3) ignoring whether words next to the fixated word were fixated or skipped. They also (4) argue that distributed processing is inconsistent with results from well-controlled experiments. In the present Reply, I address these concerns with new corpus analyses of single-fixation durations and with an experiment, leading to new empirical constraints for computational models of reading.

The Comment also raised epistemological and methodological issues that reach beyond the study of eye movements in reading. For example, at the end, the authors remind us of the at times difficult relation between the two disciplines of scientific psychology (Cronbach, 1957, 1975), namely its "experimental" and "correlational" branch. In the Reply, I argue that linear mixed-effects models are an effective way to take into account simultaneously differences between persons and differences between sentences. Such models can also test effects of many more (quasi-)experimental variables than traditional multiple-regression or ANOVA models. Moreover, the Comment highlighted the role of computational models of eye-movement control in reading. Some of them already generate complete fixation-saccade sequences for reading of isolated sentences. Their further development will be guided by answers to empirical questions about distributed lexical processing or about the relative autonomy of cognitive and oculomotor processes. They are likely to serve as a platform for general models of visual and attentional control. In fact, computational modeling may turn into the "third discipline" of scientific psychology. The current exchange offers a view on how the three "disciplines" complement each other.

The goal of research on distributed processing in reading is to determine reliable effects on fixation durations due to properties of the fixated word and those of its immediate neighbors (e.g., frequency, lexical status). In addition, there are effects of the prevailing visuo-motor characteristics (e.g., saccade amplitudes, fixation location). I will analyze all these effects and various interactions between them simultaneously in a *single* 78-predictor regression analysis. The three words cover most of the perceptual span, which is a region of roughly 4 characters to the left and at most 15 characters to the right of the point of fixation, initially mapped out by McConkie and Rayner (1975; Rayner & Bertera, 1979).

Re-Analysis of the Potsdam Sentence Corpus

Reliability and generalizability across samples

Changes in Potsdam eye-movement corpus since original publication. Kliegl et al. reported a reading-fixation corpus comprising data from 9 independent samples of readers (total N=222), varying widely in age and experimental condition who read the 144 sentences of the Potsdam Sentence Corpus while both eyes were monitored. Analyses were based on single-fixation and gaze durations identified on the same word in both eyes in first-pass reading. Binocular saccade detection was recently improved (Engbert & Mergenthaler, 2006), leading to a larger number of single fixations (N=71,097 vs. 67,260). We now also use case-sensitive word-frequency norms from the 125-million DWDS corpus (Geyken, in press; Kliegl, Geyken, Hanneforth, & Würzner, 2006) rather than those from the 5.6-million CELEX corpus (Baayen, Piepenbrock, & Rijn, 1993). All continuous predictors, like frequencies and predictabilities within readers, are centered prior to statistical analyses (Pinheiro & Bates, 2000).

From rmMRA to lme. For statistical analyses, there is a necessary switch from repeated-measures multiple regression analysis (*rmMRA*) to linear mixed-effects models (*lme*; e.g., Pinheiro & Bates, 2000), using the *lmer* program (*lme4* package; Bates & Sarkar, 2006) in the *R* system for statistical computing (R Development Core Team, 2006). The baseline model estimates 21 free parameters, that is 19 fixed effects (including the intercept) as in Kliegl et al.'s *rmMRA*, plus a random effect for readers (i.e., the variance of the intercept between readers assuming normal distribution and zero mean), and the variance of the residual error. The large ratio of 71,097 fixations to only 21 free parameters permits model estimation across all readers and separately for each of the nine independent samples.

Reliability of effects. As shown in Table 1, the results are significant and stable for independent samples comprising 18 to 33 readers, quite typical of experimental psycholinguistic research. So our results do not depend on a "huge" amount of data (Rayner et al., p. 21). They are also not due to uncontrolled heterogeneity (e.g., the inclusion of young and older readers; Rayner et al., p. 17). To the contrary, they generalize nicely across these very different groups. The collinearity of variables (e.g., between frequency and word length) or suppressor constellations (e.g., between word length and incoming saccade amplitude) *may* lead to unstable regression coefficients across studies. By this criterion, we simply do *not* have this problem. The controversial lag and successor effects of frequency and predictability are reliably observable empirical facts. Of course, the reliability of a result is only a necessary, not a sufficient condition for the validity of its theoretical explanation.

Validation with computational models and with extended regression models

There are three avenues towards establishing the validity of reliable results. First, computational models are implementations of principled accounts of reading dynamics. Thus, current models are examined to see whether they can account for small but reliable effects in reading, based on all words of sentences. Second, new variables are included in the regression model. Third, a critical result is tested with an experiment.

Validation via computational models. Computational models aspire to account for reading of all words of sentences, not only for reading of experimentally controlled target words. They must account not only for frequency effects of the fixated word, but also for frequency effects of the last word or for predictability effects of the next word. Rayner et al. reported an existence proof that E-Z Reader can generate the inverted n+1-predictability effect if an ignored covariate (i.e., a confound) is introduced into the reading material. Indeed, a confound may lead to a change in the sign of the regression coefficient; this problem exists for computational models just as it exists in regression analyses. E-Z Reader faces the challenge to predict reliable successor frequency and predictability effects of corpus data based on reading all words of sentences (see Engbert, Nuthmann, Richter, & Kliegl, 2005, for a partial success with SWIFT).

Validity problem: unobserved covariates. Computational models are useful tools to check whether theoretical reasoning is in line with the complexity of new results. To this end, we can also include new variables in the regression model. Sometimes inclusion of a new variable *may* render a formerly significant effect as insignificant; in suppressor constellations, it *may* boost its significance or even cause a change in the sign of its regression coefficient. The emphasis is on “may” because the inclusion of a new variable may also explain additional variance or change nothing at all. Rayner et al. proposed to include variables that code whether one or both of the words next to the fixated word were skipped; they also mentioned the lexical status of words, that is whether the fixated word and its neighbors are function or content words. Actually, we had considered these variables (and had done so in supplementary analyses), but left them out because frequency and lexical status are highly correlated at 0.75. Thus, the proposal raised the problem of multicollinearity, the very issue Rayner et al. found problematic in Kliegl et al.. Also with 18 predictors, we were pushing the limits of *rmMRA* because some readers contributed only around 100 fixations. Finally, none of the computational models currently uses information about the lexical status of words; they all assume that most of the variance related to lexical processing is captured with word frequency. Thus, it is common and sometimes unavoidable to start with a limited set of variables.

An extended linear mixed-effects model. Nevertheless, lexical status and skipping status are important variables, and with an *lme* model the effects of five additional moderator variables can be examined simultaneously, including also interactions between these variables and with a subset of the other variables already in the model. I also suggest how to deal with the multicollinearity of lexical status and frequency. The general framework for this analysis is shown in Figure 1. The dependent variable is the single-fixation duration on word n occurring between two saccades in the direction of reading. The neighboring words (i.e., word n-1 and word n+1) are either fixated (coded as 0) or skipped (coded as 1); this dummy coding of skipping yields tests of the effect of skipping on fixation duration. Similarly, dummy coding of lexical status of words n, n-1, and n+1 (i.e., whether they are content or function words) with three additional predictors tests the effects of these variables on fixation duration on word n. Finally, a new proposal here is to center frequencies *within* levels of lexical status for each reader to circumvent their collinearity. This recoding yields six main effects of frequencies, that is frequency nested within content words and frequency nested within function words for each of the three words n-1, n, and n+1 (see Figure 1); the cost of this recoding is loss of test statistics for interactions between frequency and lexical status. Other predictors of the baseline model (Table 1) are also included (represented with the ellipsis); they may also enter into interactions with the other variables.

The final model (including additional re-specifications and substantial pruning of non-significant interactions) comprises 79 predictors for fixed effects (incl. the intercept). The Online Supplement details the steps of building this model and how dummy coding of the variables acts as a selector of specific fixation patterns. It also contains an annotated listing of all parameter estimates and test statistics. The following examples are taken from this model and demonstrate the adequacy of current regression techniques to illuminate complex questions such as the controversial lag and successor frequency-effects in reading.

Effects of lexical status and skipping

Single-fixation duration depends on lexical status and skipping status of the word triplet. The results are summarized succinctly with two 3-factor interactions displayed in Figure 2; these interactions also qualify related main effects and simple interactions (see Online Supplement).

Skipping status of word n-1 and lexical status of words n-1 and word n. The source of the first 3-factor interaction is easily spotted (see Figure 2a; model 1: b=0.027, SE=0.010, t=2.6; the 3-factor interaction is not significant in the final model 5, but the three simple interactions still are; see Table 3 in Online Supplement). In general, fixation durations are longer after skipped words n-1—a well-known result reproduced by E-Z Reader on the assumption of a smaller preview benefit after skipping. Also shorter fixation durations in the right than the left panel are compatible with longer parafoveal preprocessing of word n during a fixation on a function word n-1 than a content word n-1—as predicted by both E-Z Reader and SWIFT. Surprisingly, significant skipping cost is observed only after function words (26 ms); they are not reliable after content words (4 ms). In addition, there is one fixation pattern opposite to the general relation: Fixations are a reliable 8 ms *shorter* if the fixated word n is a function word and the skipped word n-1 is a content word. This complex interaction is quite stable. Across the nine samples, there are only 3 of 36 possible violations of the ordinal relations depicted in Figure 2a (see Table 2).

Skipping benefit. The triple interaction is a challenge for all current models. Our interpretation of the above skipping benefit is in terms of saccadic overshoot of an intended content word. Such errors may trigger an immediate start of the next saccade program and result in shorter fixation durations (Nuthmann, Engbert & Kliegl, 2005, 2007). The immediate restart of a saccade program does not imply that the missed target is selected again (i.e., the next saccade is not necessarily corrective); rather the next saccade target is stochastically selected according to the prevailing lexical activations of all candidate words; in the SWIFT model, target selection occurs on average after 108 ms (range: 50 to 150 ms; Engbert et al., 2005). At that point in time, word n or word n+1 may often be a more competitive candidate than word n-1.

Skipping status of word n+1 and lexical status of words n+1 and word n. The second 3-factor interaction is displayed in Figure 2b (b=-0.077, SE=0.011, t=6.8). In general, fixation durations are shorter before skipped words n+1, counter to an architectural prediction of the E-Z Reader Model and also counter to simulation results of the SWIFT model (Engbert et al., 2005; Kliegl & Engbert, 2005). The results are consistent with the assumption of a quick restart of saccade programs as described above. In the present case, skipping of n+1 may be due to saccadic undershoot of the intended target word n+1. The result is very problematic for E-Z Reader because skipping requires canceling a default saccade program to word n+1 in this model. The restart of a new program, targeting word n+2, leads to enhanced fixation durations prior to skipped words.[2]

Skipping cost. One pattern in Figure 2b (left panel) appears to be consistent with the E-Z Reader prediction: Fixation durations on word n are longer if the fixated word n is a function word and if the word to be skipped is a content word. Again, there is a high stability of this result across the nine samples (see Table 3, also for percentages of fixations in the cells). This pattern, however, is also a problem for the E-Z Reader model, because saccade programs are cancelled by an attention shift from word *n*+1 to word *n*+2, which depends on an easy word *n*+1; indeed this mechanism allows E-Z Reader to skip function words more frequently than content words. In contrast, here fixation durations are longest prior to skipped difficult content words *n*+1. There is a straightforward interpretation of this particular skipping cost. Readers fixate a function word, say a determiner, but they process the noun or adjective to the right of it, that is, there is genuine lexical processing of parafoveal words in the perceptual span (i.e., Radach's, 1996, "word-group" hypothesis). This result accounts for the interaction between word length of word n and frequency of word n+1, reported by Kliegl et al., Figure 4c, as well as Kennedy and Pynte (2005). Interestingly, this effect is not present for gaze durations (see Online Supplement, Figure 3b, left panel). E-Z Reader must assume that these fixations result from failed skipping of the function word (i.e., an undershoot of the intended content word) and that word recognition proceeds as if word *n*+1 were fixated (i.e., causing normal lexical access). The question then is: Why should "mislocated" fixations occur only on function words prior to content words that are subsequently skipped? And why are they shorter?

Experimental evidence. Critical corpus-analytic results must be followed up with experiments. In a recent one, modeled after Rayner, Juhasz, and Brown (2007), Kliegl, Risse, and Laubrock (2007) presented word n+2 normally or as a random letter string until the eyes crossed from word n to word n+1. In this experiment, word n+1 (after the boundary, but always visible) was either a three-letter function word or a three-letter content word; word n+2 was almost always a content word. I analyzed log of first-fixation durations on three-letter words n+1, contingent on skipping status of word n+2. In agreement with the corpus analysis, there is skipping "cost" in the function-word condition (fixate n+2: 201 ms; skip n+2: 222 ms) and skipping "benefit" in the content-word condition (fixate n+2: 195 ms; skip n+2: 190 ms). In addition, denial of preview of word n+2 increased the skipping "cost" measured on word n+1(with preview: 16 ms, without preview: 27 ms); the statistic for the three-factor interaction of lexical status, skipping status, and preview was b=0.159, SE =0.058, t=2.7. The "inconsistency" between corpus-analytic and experimental evidence identified by Rayner et al. has been reduced with this experiment.

Summary. In perspective, the distinction between function and content words needs to be addressed with experiments and with computational modeling. There is statistically significant evidence against E-Z Reader's saccade cancellation mechanism, but SWIFT also does not adequately reconstruct the pattern of results, as already described and discussed in Engbert et al. (2005). In the latter model, the delay of saccade programs (i.e., modulation of single-fixation durations) is linked strictly to the lexical activation of the fixated word. Possibly, other words in the perceptual span exert a much stronger control than currently allowed for.

Effect of frequency of content word n-1

Is the n-1-CW-frequency effect due to lexical status and skipping status of word n-1? Given the coding of Figure 1, the n-1-CW-frequency main effect represents only fixation patterns where the eyes fixated on three successive content words, including a single fixation on word n (i.e., three content words, no skipping). The n-1-CW-frequency effect for this pattern, typical of psycholinguistic experiments, is b=-0.032 (SE=0.002, t=-14). Thus, fixation duration on content word n strongly depend on the frequency of content word n-1, even after statistical control of the other word properties, the incoming and outgoing saccade amplitudes, and the effect of fixation position within word.

Estimates of interactions with n-1-CW-frequency inform about fixation patterns that differ from the baseline pattern (see Online Supplement for details). For example, if word n-1 is skipped, the n-1-CW-frequency effect is Δb=+0.039, SE=0.004, t=11. This value must be added to the baseline coefficient, yielding a b= -0.032+0.039=0.007, which is no longer significant. Thus, apparently the frequency of skipped content words n-1 does not contribute to the n-1-CW-frequency effect. In contrast, for function words n, the n-1-CW-frequency effect is significantly reduced (Δb=+0.019, SE=0.003, t=5.6), but still significantly negative (b= -0.032+0.019=-0.013). Thus, the n-1-CW-frequency effect is weaker in the presence of function words n and possibly even absent for skipped content words n-1. The Online Supplement lists the reliable interactions for the various fixation patterns, also those relating to the n-1-FW-frequency effect.

Effect of frequency of word n+1

Similarly, the n+1-CW-frequency effect cannot be reduced to the three-factor interaction displayed in Figure 2b. First, the effect is reliable for the baseline pattern (i.e., three content words, no skipping; b=-0.0035, SE=0.0016, t=-2.2). Second, the effect is magnified if word n+1 is skipped (Δb=-0.018, SE=0.004, t=-4.5). Adding this value to the baseline term yields a coefficient b=-0.004+(-0.018)=-0.022. Thus, skipping of words n+1 is a primary source of the n+1-CW-frequency effect, in agreement with the results suggesting parafoveal processing of content words.

Effect of predictability of word n+1

The n+1-predictability effect is b=0.0036 (*SE*= 0.0016, *t*=2.3). This effect is significantly stronger for fixations on function words n (Δb=+0.009, *SE*=0.003, *t*=3.0, model 5) or if word n+1 is a function word (Δb=+0.008, *SE*=0.002, *t*=3.6, model 5). Both effects are in agreement with the retrieval account. These two n+1-predictability effects are underadditive: if both word n and word n+1 are function words, then Δb=-0.031, *SE*=0.004, *t*=-6.1), yielding a negative (!) net n+1-predictability effect of b=0.004+0.009 +0.008-.031=-0.1. There is no reliable statistical evidence that skipping status of word n+1 moderates this effect. As function words are rarely manipulated, previous experiments may easily have missed the effect.

Limitations

The present model specification is not the only conceivable one. For example, possible interactions with incoming and outgoing saccade amplitudes are ignored. Further, reader-level variables such as skipping rate or mean saccade length may be relevant as well; a different coding scheme and inclusion of other predictors (e.g., orthographic word properties) may foreground other effects and lead to an alternative account of the results. This is “normal science” and this is not different from the situation faced by researchers employing experimental designs; we cannot be sure that all relevant variables are under statistical or experimental control—some influences may still have to be discovered. I picked only a few topics that I perceived as relevant for theoretical controversies and the Comment. The present regression analysis demonstrates how Rayner et al.’s concerns, including those not explicitly addressed here, can be constructively dealt with by bringing new variables into the model.

(Quasi-)Experimental and Corpus-Analytic Research

Rayner et al. categorized reading research as either experimental or correlational. Well-controlled experiments are defined by a random assignment of units to levels of the independent variables. Words (like people) can be randomly assigned to levels of a genuine experimental variable such as preview/no-preview, high/low monitor contrast. Words (like people) cannot be randomly assigned to levels of dimensions inherent to them (words: frequency, predictability, etc.; persons: gender, age, etc.). Only random assignment allows us to draw inferences about the causal role of an experimental factor, within limits of false-positive errors. Quasi-experimental variables afford only *correlational* inference and no amount of control will convert them into experimental factors. Moreover, interactions with such a variable (e.g., frequency by preview; age by preview) also yield only correlational results.

Lag frequency, parafoveal predictability, etc. are quasi-experimental variables and, therefore, they do not allow any stronger causal inferences in experiments than in corpus analyses; they always yield correlational evidence. Frequency effects may be due to correlations with sublexical properties of word beginnings, but this applies to all frequency effects, irrespective of whether they are measured in an orthogonal design or in a corpus analysis and irrespective of whether they are determined with *ANOVA*, *rmMRA*, or *lme*.

Rayner et al. presumably grant that most psycholinguistic research is correlational in this strict sense. I agree with their concerns about interpreting regression coefficients of correlated predictors—who would deny the beauty of an orthogonal (quasi-)experimental design? Typically, regression analysis ignores the shared variance in the dependent variable for significance tests, and this increases the danger of a false-negative error (i.e., reduces the power of the test). In other words, with correlated predictors one is in a very weak position to argue the null hypothesis. Our “problem” (according to Rayner et al.), however, is not so much the *absence*, but the *presence* and *size* of significant lag and successor effects. To the contrary, I suspect that some of these inconsistencies relate to low statistical power for participants or items in some of the experiments. I also have shown for the case of lexical status and frequency how nesting of a factor can help to circumvent a serious problem of collinearity.

Generalizability of words and issues of data selection

Rayner et al. perceived it as “a problem … that [we] include all words” (p. 17), because this way short words dominate. This argument can be turned around because experiments carry their own biases. For example, one source of inconsistency in results between corpus analysis and experiments may be the restricted range of word frequency of experimental target words with negative consequences for statistical power. Such a restriction may be necessary if specific aspects of word recognition during reading are the focus of the research. It is equally legitimate and necessary to look at the complete range of word properties if the interest is in the dynamics of reading. Given that reading for the most part involves short and highly frequent words, computational models of reading should respect regularities related to them.

There are differences between experiments and corpus analysis relating to data selection. In Kliegl et al., we separately examined single-fixation and multiple-fixation cases (dominated by long words) and did not hide a "bias" in the respective analyses. Our Table 1 provides information about characteristics of eye movements and words selected for the various types of analyses. So researchers inclined to compare their results with ours can do so. We reported on similarities and differences of lag/successor effects between single-fixation and gaze durations; new ones are presented in the Online Supplement, including descriptive statistics and the result of the *lme* analysis of gaze durations. What is needed next is an integrative data-analytic framework to test explicitly differences of distributed processing in single-fixation and multiple-fixation cases.

Conclusion

The ultimate goal of reading research is to understand how we read normal text. An immediate goal is to understand what information becomes available when from where in the perceptual span. A wide range of techniques is available for studying this process: (1) experiments on visual, oculomotor, language-related, and neural correlates of reading, (2) computational models of eye-movement control, and—last not least—(3) regression analyses of eye movements during reading of texts and sentences. Experiments and computational models are indispensable tools for the validation of theoretical principles; they are not ends in themselves. Regression analysis is not only also a legitimate tool, sometimes it may be the only way forward and even set the pace for experiments and computational models, if we want to understand the full complexity of the multidimensional space being studied.

References

Baayen, R.H., Piepenbrock, R., & Rijn, H. van (1993). *The CELEX Lexical Database (Release 1)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bates, D., & Sarkar, D. (2006). *lme4: Linear mixed-effect models using S4 classes.* R package version 0.995-2.

Blouin, D.C., & Riopelle, A.J. (2005). On confidence intervals for within-subject designs. *Psychological Methods, 10,* 397-412.

Cronbach, L.J. (1957). The two disciplines of psychology. *American Psychologist, 12,* 671-684.

Cronbach, L.J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116-127.

Drieghe, D., Rayner,. K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 954-969.

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal slip. *Proceedings of the National Academy of Sciences, 103,* 7192-7197.

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

Geyken, A. (in press). The DWDS-Corpus: A reference corpus for the German language of the 20th century. In C. Fellbaum (ed.), *Collocations and idioms: linguistic, lexicographic, and computational aspects.* London: Continuum Press.

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

Kliegl, R., & Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychonomic Bulletin & Review, 12,* 132-138.

Kliegl, R., Geyken, A., Hanneforth, T., & Würzner, K., (2006). *Corpus matters: A comparison of German DWDS and CELEX lexical and sublexical frequency norms for the prediction of reading fixations.* Manuscript.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 133,* 12-35.

Kliegl, R., Risse, S., & Laubrock, J. (2007). *Preview benefit and parafoveal-on-foveal effects from word n+2.* Manuscript accepted pending revision.

McConkie, G.W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578-586.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research, 45,* 2201-2217.

Nuthmann, A., Engbert, R., & Kliegl, R. (2007). The IOVP effect in mindless reading: Experiment and modeling, *Vision Research, 47,* xxx-xxx.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-Plus.* New York: Springer.

Plourde, C.E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology, 51,* 181-194.

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.

R Development Core Team (2006). *R: A language and environment for statistical computing*. (version 2.3.1). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Radach, R. (1996). *Blickbewegungen beim Lesen [Eye movements during reading].* Muenster: Waxmann.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K., & Bertera, J.H. (1979). Reading without a fovea. *Science*, *206*, 468-469.

Rayner, K., Juhasz, B.J., & Brown, S.J. (2007). Do readers obtain preview benefit from word n+2? A test of serial attention shift vs. distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance.*

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T.J., & Reichle, E.D. (2007). Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert (2006). *Journal of Experimental Psychology: General, 134,* xxx-xxx.

Rayner, K., Reichle, E.D., & Pollatsek, A. (2005). Eye movement control in reading and the E-Z Reader model. In G. Underwood (ed.), *Cognitive processes in eye guidance.* (pp. 131-162). Oxford: OUP.

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.

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.

Author Note

I thank Ralf Engbert and Antje Nuthmann for entrusting me with a single-author reply. I am grateful to them as well as to Harald Baayen, Sachiko Kinoshita, Jochen Laubrock, Mike Masson, Keith Rayner, and Shravan Vasishth for helpful comments. This research was supported by grant KL 955/6-6 from Deutsche Forschungsgemeinschaft. Correspondence: Reinhold Kliegl, Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany, email: reinhold.kliegl@uni-potsdam.de.

Footnotes

1. The intended meaning of “natural reading” in Kliegl et al. was “reading of natural texts” (Rayner et al., in press, p. 30), contrasting analyses of fixations from texts or sentences with varying difficulty and many short words, on the one hand, with analyses of fixations on *a priori* specified experimentally manipulated target words, often only content words, on the other hand; we certainly do not think of the latter as unnatural reading. Erratum: We cited Plourde and Besner (1997) as support of a frequency(*n*) by visual-degradation interaction, although the point of their paper was to demonstrate an additive relation.

2. Drieghe, Rayner, and Pollatsek (2005, p. 956; also Rayner, Reichle, & Pollatsek, 2005, p. 141) discount skipping benefit as “correlational” and speculate that it is due to fluctuations of attention or text difficulty; Pollatsek et al. (2006, simulation A4, p. 46) ignore them. Their position may be valid for early reports, including, for example, Schilling et al. (1998). In Kliegl and Engbert (2005), however, we employed matching and resampling procedures to guard exactly against these possible confounds and a few more (e.g., fixation position within words). Of course, this does not render our data as experimentally well controlled, but they certainly are statistically well controlled—barring new evidence to the contrary.

Table 1. Parameter estimates for fixed effects of random-intercept *lme* models (all + 9 samples)

Sample	all	1	2	3	4	5	6	7	8	9
(intercept)	212	216	201	222	221	200	210	212	211	222
Word n										
frequency (frq)	-5.0	-7.0	-2.7	-6.2	-8.2	-3.4	-4.9	-5.6	-4.1	-3.7
predictability (prd)	-4.3	-4.1	-3.4	-6.8	-2.8	-3.7	-4.4	-6.0	-3.5	-6.4
1/length	42	**10**	40	**17**	42	28	80	62	49	51
Word n-1										
frequency	-6.6	-6.6	-5.5	-7.5	-6.8	-5.6	-6.9	-7.4	-6.0	-8.2
predictability	-1.2	-1.9	-1.5	**-1.1**	-2.2	-1.4	**0.7**	-2.2	**-0.2**	**-0.3**
1/length	35	53	37	32	45	24	29	27	**15**	45
Word n+1										
frequency	-3.9	-2.6	-4.8	-3.2	-2.3	-4.1	-4.1	-5.3	-3.2	-5.0
predictability	2.7	2.3	3.2	1.4	2.6	2.0	3.4	2.7	3.2	2.6
1/length	**2.7**	**13**	**12**	**6**	**-13**	**-4**	**-3**	**10**	**-4**	**15**
Viewing position										
last sacc. ampl.	4.8	5.5	4.5	4.8	4.9	4.1	5.1	4.4	4.4	5.4
pos in word	-26	-16	-31	**0.8**	-31	-20	-36	-35	-27	-23
pos^2 in word	-72	-53	-94	**4.1**	-78	-53	-89	-85	-77	-88
next sacc. ampl	1.2	**0.0**	0.8	**0.6**	2.1	1.2	2.3	**0.8**	**0.8**	2.7
Interactions										
(frq n)/(lgth n)	22	**33**	**11**	44	27	16	22	21	16	20
(frq n)*(frq n-1)	2.1	**0.8**	3.0	**0.9**	1.7	2.0	2.1	2.0	3.0	2.2
(frq n)*(frq n+1)	**0**	**0**	**-0.3**	**-0.2**	**0.1**	**0.4**	**-0.1**	**0.2**	**-0.4**	**0**
(frq n+1)/(lgth n)	11	**12**	**9**	**16**	**5**	**5**	**14**	**3**	20	20
(prd n+1)/(lgth n)	-14	-20	**-11**	-25	-18	**-1**	-17	-18	-15	**-13**
Reader SD	30	29	32	22	28	30	32	25	26	22
Residual SD	57	55	56	60	60	51	57	57	57	64
N fixations	71097	8814	11312	4938	9885	9492	8373	5765	6593	5925
N readers	222	32	33	18	29	27	24	19	22	18

Note. **Non**-significant coefficients ($b < 2$ SE) are set in **bold**; models were fit by restricted maximum likelihood.

Table 2. Mean fixation durations on word *n* after fixated and after skipped words *n*-1, broken down by lexical status of word n and word n-1.

word n	content word (CW)				function word (FW)			
word n-1	CW		FW		CW		FW	
word n-1	fx	sk	fx	sk	fx	sk	fx	sk
sample 1	222	234	191	224	*208*	212	188	213
2	*205*	*203*	176	200	207	192	194	196
3	234	242	201	225	222	220	206	218
4	231	239	200	227	224	215	200	223
5	*206*	*202*	180	195	204	191	193	209
6	220	220	187	214	226	215	201	210
7	225	228	187	213	221	210	202	208
8	218	227	185	211	212	209	204	217
9	231	231	199	218	225	219	216	235
M	220	224	188	214	**216**	**208**	200	213
SD	67	65	61	60	71	64	68	63
N	24401	4187	16476	7019	9970	3454	3183	2407
%	34	6	23	10	14	5	4	3

Note. fx=fixated, sk=skipped before fixation on word *n*. Key "skipping benefit" is printed in bold; three numerical violations of dominant skipping pattern in conditions set in *italics*.

Table 3. Mean fixation durations on word *n* before fixated and before skipped words *n*+1, broken down by lexical status of word n and word n+1.

word n	content word (CW)				function word (FW)			
word n+1	CW		FW		CW		FW	
word n+1	fx	sk	fx	sk	fx	sk	fx	sk
sample 1	217	207	215	215	203	222	*205*	*214*
2	197	188	200	186	199	212	202	198
3	225	218	231	214	218	232	*214*	*217*
4	222	205	226	215	216	234	220	214
5	199	183	200	187	202	204	199	189
6	210	197	216	200	215	230	225	203
7	213	199	213	208	215	226	214	191
8	208	192	210	204	211	215	207	205
9	220	211	221	213	221	245	223	222
M	211	198	214	203	**210**	**222**	212	205
SD	67	62	66	62	68	75	67	61
N	25309	3046	13501	10227	12263	2234	2585	1932
%	36	4	19	14	17	3	4	3

Note. fx=fixated, sk=skipped after fixation on word *n*. Key "skipping cost" is printed in bold; two numerical violations of dominant skipping pattern in conditions set in *italics*.

Figure Caption

Figure 1. Distributed processing framework for possible influences on single-fixation duration (●) on word n; arrows represent possible incoming (left) or outgoing (right) saccades (i.e., word n-1 and word n+1 are fixated or skipped); lexical status = content (CW) or function (FW) word; "|" read as "given", that is, for example, "frequency given content word n-1" (otherwise frequency is set to zero). Thus, frequencies are specified as nested within lexical status; predictability is specified as crossed with lexical status. (…) = additional predictors (e.g., Table 1, for complete list see Online Supplement).

Figure 2. Two three-factor interactions for single-fixation durations. 99% confidence intervals of *lme* model are 3 ms (Blouin & Riopelle, 2005). (a) Interaction of skipping status of word n-1, lexical status of word n-1, and lexical status of word n. Skipping "benefit" is observed only for fixations on a function word following a skipped content word n-1. (b) Interaction of skipping status of word n+1, lexical status of word n+1, and lexical status of word n. Skipping "cost" is observed only for fixations on functions word n if a subsequent content word n+1 is skipped. Skipping "benefit" is explained with immediate restart of saccade program after saccadic errors (saccadic overshoot in Figure 2a; saccadic undershoot in Figure 2b); skipping "cost" is explained as lack of preview (Figure 2a) and parafoveal lexical processing of content word n+1, given a fixation on a function word n (Figure 2b).

Figure 1

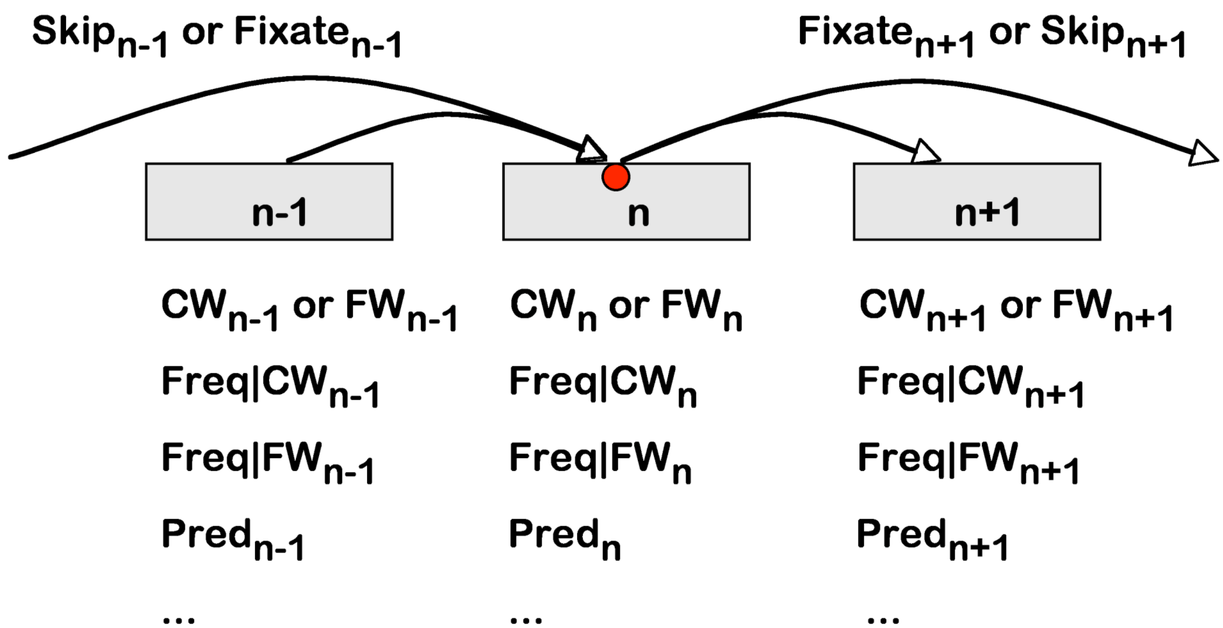
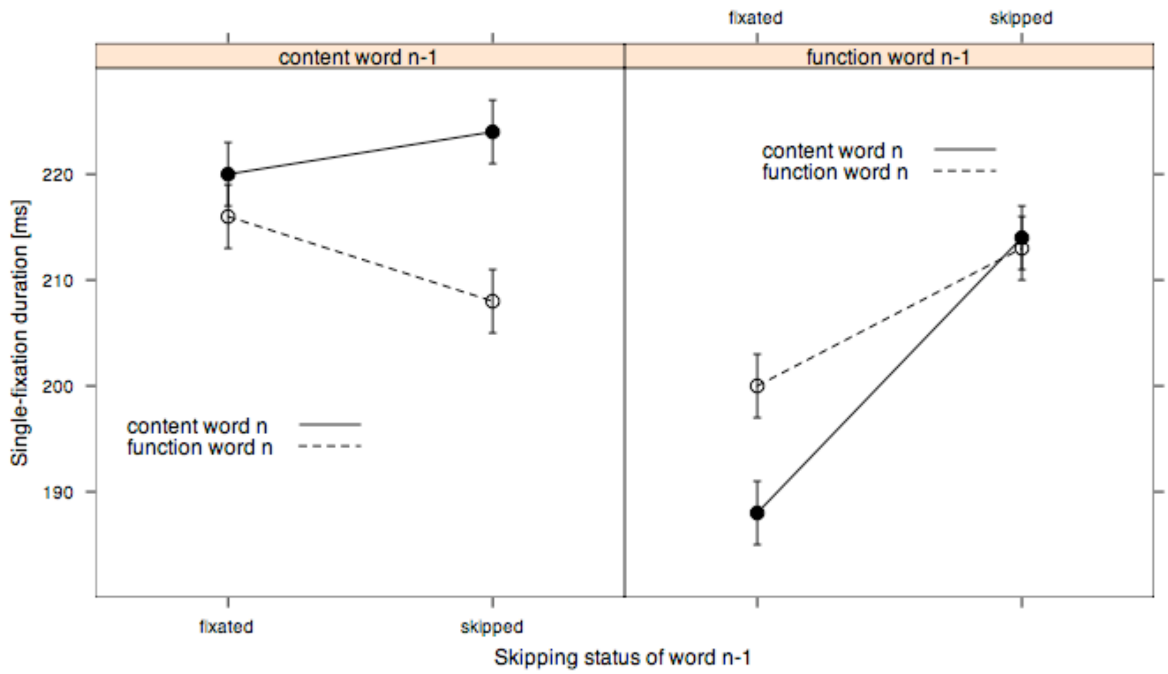

Skip$_{n-1}$ or Fixate$_{n-1}$
Fixate$_{n+1}$ or Skip$_{n+1}$
n-1
n
n+1
CW$_{n-1}$ or FW$_{n-1}$
Freq|CW$_{n-1}$
Freq|FW$_{n-1}$
Pred$_{n-1}$
...
CW$_{n}$ or FW$_{n}$
Freq|CW$_{n}$
Freq|FW$_{n}$
Pred$_{n}$
...
CW$_{n+1}$ or FW$_{n+1}$
Freq|CW$_{n+1}$
Freq|FW$_{n+1}$
Pred$_{n+1}$
...

Figure 2

(a) Interaction of lexical status n, lexical status n-1, and skipping status n-1



(b) Interaction of lexical status n, lexical status n+1, and skipping status n+1