University of Cincinnati

Date: 11/3/2011

I, Sebastian Wallot, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Psychology.

It is entitled:
The role of reading fluency, text difficulty and prior knowledge in complex reading tasks.

Student's name: Sebastian Wallot

This work and its defense approved by:

Committee chair: Guy Van Orden, PhD
Committee member: John Holden, PhD
Committee member: Beth O'Brien, PhD
Committee member: Michael Richardson, PhD

2024
The role of reading fluency, text difficulty and prior knowledge in complex reading tasks.

A dissertation submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

in the Department of Psychology
of the College of Arts and Sciences
by

Sebastian Wallot

Dipl. Psych., University of Trier
November 2011

Committee Chair: Guy Van Orden, Ph.D.
Committee: Beth O'Brien, Ph.D., John Holden, Ph.D., & Michael Richardson, Ph.D.
Abstract

Despite the fact that reading is one of the most frequently studied topics in psychology, cognitive science, and educational science, research on text reading has made only little progress in the past 100 years. This is true for the more basic research on reading as a cognitive activity, where materials that resemble natural texts are usually neglected in favor of tasks that employ only individual words or sentences. It is just as true for applied research, where studies either follow the lead of the ‘few-words-approach’ or focus exclusively on the outcome of reading performance – such as memory or comprehension of a read passage. In this latter case, the reading activity itself that brings about a certain memory or comprehension result remains in a black box. The work presented in this dissertation tries to bridge this gap. In particular, the aim is to expand the work of Wallot and Van Orden (2011a, b) on the application of complexity metrics to evaluate the process of reading in complex text reading tasks.

Complexity metrics, which quantify the degree of stability, complexity, and interconnectedness of performance, are employed to expand the findings on reading fluency in self-paced reading by Wallot and Van Orden (2011a, b) to the level of eye-movements during reading. While reading performance during self-paced reading is estimated by the intervals between key-presses that the reader employs to reveal each new word or sentence of a text, eye-movements are considered to be a more fine-grained measure of the reading process (Rayner, 1978). Hence, study 1 of this dissertation is an attempt to replicate the findings of Wallot and Van Orden (2011a, b), where the time evolutions of the performance of more fluent readers showed higher
stability and greater commonality compared to less fluent readers, and the stability of
and commonality between less fluent readers’ performances increased with repeated
reading of the same text. Study 2 gathers first evidence for what distinguishes the
process of reading of an easy text from the process of reading of a difficult text.

While the results of self-paced reading in study 1 do not replicate all of the effects
reported by Wallot and Van Orden (2011a, b), they confirm that the performances of
fluent readers share more dynamic structure than the performance of less fluent readers
and that the used non-linear methods can successfully distinguish between these two
reader groups. The results of eye-movement dynamics show effects similar to those of
key-press intervals and also replicate the gain in stability of less fluent readers with re-
reading. This seems to suggest that eye-movements are indeed a more sensitive
measure that can pick out differences in performance that are not picked out in key-
press intervals.

The results of study 2 suggest that high text difficulty constrains the dynamics of
reading measures in key-press intervals and eye-movements alike, since key-press
intervals and eye-movements during reading show greater stability during difficult text
reading compared to easy text reading.

In conjunction, the two studies provide strong evidence for the utility of
complexity metrics to quantify reading performance. Furthermore, the studies present
evidence that key-press and eye-movement measures of reading seem to be driven by
similar organizational principles, following similar dynamics during reading tasks.
However, the results of the two studies seem to be contradictory with regard to the role
of dynamic stability, which seems to indicate higher reading skill in study 1, but more
difficult reading in study 2. Different possible interpretations of these results are discussed. It is concluded that the observed reading performance can be seen as a synergy between text and reader, and that the stability and complexity of this synergy holds information about the overall properties of reading tasks.
Acknowledgements

I would like to thank a few persons who went with me through (or steered me through) the three years of graduate school at the University of Cincinnati and made these years a scientifically and personally a rich and pleasant experience. The order of these acknowledgements can be seen as meaningful in that regard, but with one exception: The pole position is does not belong to the ‘biggest contributor,’ but is--in keeping with scientific world view that I have adopted during the past three years--qualitatively different from the others and cannot be brought in a simple relation to them. It belongs to Anna.


vi
Kisses.

Of course, the first on this list is Guy. This is in good keeping with the tradition of mowing your advisor’s lawn to avoid getting kicked out of grade school, or putting his name on every paper that you publish, even if he (or she, or course) doesn’t even know of its existence. Mainly, however, it is because I would not have come here without you, Guy, in the first place, and there are reasons for that: The first being, that you--as I suspect now, on purpose--send me corrupted files via email that should have been reprints of your research papers so often, that by the time the actual document made it into my mailbox, I could not help but apply for grad school with you as my advisor. The second (and third) being, that you did not only support me generously in word and deed, but also that we have become friends over those past three years, a friendship that I treasure. The fourth being, that what I know today, and what will (hopefully) make me a good scientist, I have learned from, through, and with you. Of course, the last statement might have beard somebody else’s name if my life would have wound down different paths, but it didn’t, and as far as we know, for us, organisms, the evolution of reality is not made to be unwound. Thank you. I am grateful for all that you did for me. And I hope that you will be with me for many more years, despite the odds.

The second person I want to acknowledge in this dissertation is Nikita Kuznetsov. You are not only one of the most able and serious persons that I have come to know in the past three years. No, you are also a person of substantial moral integrity, not only always weighting all facts carefully in matters scientific, but also social and interpersonal. That was inspiring for somebody like me, who--depending on the actual
parametrization with regard to caffeine/sleep/general mania--does really talk too much too often. Moreover, I have learned with you and from you as much as from anyone. And I hope this will not change for the years to come.

The last person I want to acknowledge is Mike Richardson. In some sense you came completely out of nowhere, claiming an office space that lay on my daily route to the PMD lab (which I travelled often for procrastination). Perhaps this is the reason why I had to--inevitably--steal a couple of minutes of your life’s time every day to gossip, talk about kids or research. Thanks for letting me. As time went by, and I again didn’t notice the point at which it happened, you somehow became some sort of second advisor, sharing knowledge and curiosity with me. Coming in in the morning and seeing your office door open has always made my day a little better, at least never worse. Thank you for that.

Of course, there are many other people who contributed to making the ‘grad-school-experience’ a better and friendlier one. But to the four persons explicitly mentioned here, I want to say: I salute you! It just wouldn’t have been the same.

Sebastian Wallot

Cincinnati in November 2011
# Table of Contents

List of Figures ................................................................. xii

CHAPTER 1: Text Reading .................................................... 1

CHAPTER 2: Nonlinear Methods ............................................. 13

- Fractal Analysis ......................................................... 13
- Monofractal Analysis ..................................................... 16
- Multifractal Analysis ..................................................... 18
- Recurrence Quantification Analysis .................................. 20
- Cross-Recurrence Analysis ............................................. 25
- What Can and Do Nonlinear Analysis Techniques Offer? ........ 27

CHAPTER 3: Study 1 ............................................................... 28

- Method ................................................................. 30
  - Design and Participants .................................................. 30
  - Apparatus and Stimuli .................................................... 31
  - Procedure .............................................................. 33
  - Data Preparation ....................................................... 34
- Results of Key-Presses .................................................... 37
  - Averages and Standard Deviations .................................... 37
  - Mono- and Multifractal Analysis ..................................... 38
  - Recurrence Quantification Analysis .................................. 40
  - Cross-Recurrence Quantification Analysis .......................... 42
- Discussion of Key-Presses ................................................ 44
Results of Fixations and Saccades ................................................................. 46
Averages and Standard Deviations .............................................................. 46
Mono- and Multifractal Analysis ................................................................. 49
Recurrence Quantification Analysis ............................................................ 51
Cross-Recurrence Quantification Analysis ............................................... 53
Discussion of Fixations and Saccades ......................................................... 56
Conclusion .................................................................................................. 60
CHAPTER 4: Study 2 ..................................................................................... 64
Method ........................................................................................................ 65
Design and Participants ............................................................................... 65
Apparatus and Stimuli .................................................................................. 66
Procedure ................................................................................................... 68
Data Preparation .......................................................................................... 68
Results of Key-Presses ............................................................................... 68
Averages and Standard Deviations .............................................................. 68
Mono- and Multifractal Analysis ................................................................. 68
Recurrence Quantification Analysis ............................................................ 71
Cross-Recurrence Quantification Analysis ............................................... 72
Discussion of Key-Presses ......................................................................... 73
Results of Fixations and Saccades ............................................................... 79
Averages and Standard Deviations .............................................................. 79
Mono- and Multifractal Analysis ................................................................. 81
Recurrence Quantification Analysis ............................................................ 83
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-Recurrence Quantification Analysis</td>
<td>84</td>
</tr>
<tr>
<td>Discussion of Fixations and Saccades</td>
<td>87</td>
</tr>
<tr>
<td>Conclusion</td>
<td>88</td>
</tr>
<tr>
<td>CHAPTER 5: General Discussion</td>
<td>91</td>
</tr>
<tr>
<td>Reading Skill, Text Difficulty, and Prior Knowledge</td>
<td>92</td>
</tr>
<tr>
<td>Entrainment as the Core-Concept of Text Reading</td>
<td>97</td>
</tr>
<tr>
<td>Eye-Movement and Key-Press Measures of Reading Performance</td>
<td>99</td>
</tr>
<tr>
<td>Conclusion and Outlook</td>
<td>100</td>
</tr>
<tr>
<td>References</td>
<td>102</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1. Illustration of the self-paced reading task. ...

Figure 2.1. Top: The time series of one participant’s word reading times. Middle: The same time series, but with extreme outliers removed and normalized is subjected to spectral power density analysis (bottom) where sine and cosine waves are fitted to the data to estimate the power at different frequencies. The Power spectrum is plotted on log-log axis to investigate regions of scaling in the time series.

Figure 2.2. From a one dimensional time-series to a three-dimensional reconstructed phase-space to a recurrence plot. The top row illustrates how a time-series plotted against itself at time t4, t8, and t12 yields coordinates for a three-dimensional phase-space portrait of that time-series (i.e., x = t4, y = t8, z = t12). The bottom row illustrates how a neighborhood (indicated by the red circle) in the phase-space gives rise to one or more points in the recurrence plot.

Figure 3.1. Average sentence reading times decrease with repeated reading (left). Average standard deviations of sentence reading times decreases with repeated reading (right). Averages are bracketed by the standard errors of the means.

Figure 3.2. Neither alpha exponents (left) nor multifractal spectrum widths (right) differ reliably as a function of reader group or repeated reading. Averages are bracketed by the standard errors of the means.

Figure 3.3. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) differ reliably as a function of reader group or repeated reading. Averages are bracketed by the standard errors of the means.

Figure 3.4. Shared %REC (upper-left), %DET (upper-right), %LAM (lower-left), and ENTR (lower-right) among graduate readers are reliably higher compared to undergraduate readers. Averages are bracketed by the standard errors of the means.

Figure 3.5. Average duration of fixations decreases reliably with repeated reading (upper-left). Average standard deviation of fixations decreases reliably with repeated reading (upper-right). Average duration of saccades generally increased with repeated reading, but the increased was smaller for graduate readers compared to undergraduate readers (lower-left). Average standard deviation of saccades did not differ reliably as a function of reader group or repeated reading (lower-right). Averages are bracketed by the standard errors of the means.
Figure 3.6. The number of fixations dropped reliably with repeated reading (left). The number of saccades dropped reliably with repeated reading (right). Averages are bracketed by the standard errors of the means. ....... .............. .............. .............. 48

Figure 4.3. %REC (upper-left), %DET (upper-right), %LAM (lower-left), and ENTR (lower-right) of sentence reading times were reliably higher for the difficult text compared to the easy text. Averages are bracketed by the standard errors of the means.

Figure 4.4. Shared %REC (upper-left), %DET (upper-right), %LAM (lower-left), and ENTR (lower-right) of sentence reading times were reliably higher for the difficult text compared to the easy text, but this difference was less pronounced when the first part of the text was skipped. Averages are bracketed by the standard errors of the means.

Figure 4.5. Correlations of sentence length with key-press intervals before (upper-left) and after (upper-right) parceling out the scaling relation between power and frequency observed in spectral analysis. Alpha-values for key-press intervals before (lower-left) and after (lower-right) parceling out the correlation between sentence length and key-press intervals. Averages are bracketed by the standard errors of the means.

Figure 4.6. Average duration of fixations was reliably longer for the difficult text compared to the easy text (upper-left). Average standard deviation of fixations was reliably greater for the difficult text compared to the easy text (upper-right). Average duration of saccades (lower-left) and average standard deviation of saccades (lower-right) did not differ as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.

Figure 4.7. The number of fixations (left) and saccades (right) was higher for the difficult text compared to the easy text. Averages are bracketed by the standard errors of the means.

Figure 4.8. Alpha exponents were higher for the difficult text compared to the easy text for fixations (upper-left) as well as for saccades (lower-left). Multifractal spectrum widths did not reliably differ as a function of text difficulty or the presence or absence of the first 30% of the text, neither for fixations (upper-right) nor saccades (lower-right). Averages are bracketed by the standard errors of the means.

Figure 4.9. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) of fixations differed reliably as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.

Figure 4.10. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) of saccades differed reliably as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.
Figure 4.11. Shared %REC was higher for the difficult text compared to the easy text (upper-left). Shared %DET was higher for the difficult text compared to the easy text, and shared %DET was higher when the whole text was presented compared to when the first 30% of the text were skipped. Furthermore, when the first 30% of the text were skipped, then shared %DET was disproportionally lower for the easy text compared to the difficult text (upper-right). Shared %LAM was higher for the difficult text compared to the easy text (lower-left). Shared ENTR was higher for the difficult text compared to the easy text, and shared ENTR was higher when the whole text was presented compared to when the first 30% of the text were skipped (lower-right). Averages are bracketed by the standard errors of the means.

Figure 4.12. Shared %LAM of saccades was higher when then whole text was presented compared to when the first 30% of the text were skipped. Neither shared %REC, nor shared %DET or shared ENTR differed reliably as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.
CHAPTER 1:
Text Reading

Reading, together with writing, is not only one of the hallmark activities that distinguishes humans from other creatures, but is also one of the--if not the--most studied cognitive skill in the laboratories of psychologists and cognitive scientists. There is not much that has not been tried in the investigation of the perception of written language, from scrambled word reading (Rawlinson, 1976), to non-word-symbol insertion into texts (Epelboim, Rooth, Ashkenazy, Taleghani, & Steinman, 1997) or the investigation of cross-modal effect of spoken on written language perception (Luce & Cluff, 1998). However, one fact catches the eye: Reading is almost always studied either in terms of single words (standard word naming or lexical decision tasks), word pairs (priming tasks), or single sentences, a hand full of sentences at most.

In ‘The Psychology of Reading’, Rayner and Pollatsek (1994) survey the approaches that cognitive scientists have taken to investigate the process of text reading, and most of the experimental setups use no more than two sentences, perhaps holding at about twenty words altogether. The same is true for an overview article by Clifton and Duffy (2001), where studies of ‘text’ reading encompassed eight sentences at the most. The layman might be puzzled by this sparsity, but this very sparsity has been adopted by psycholinguists and other reading researchers for good reason: consecutive presentations of as few as two words in a row already result in all kinds of complicated carry-over effects.
For example, a key-press response in a simple reading task, to indicate that the word *pepper* is indeed a word (with respect to English spelling) will be about 48 ms faster (on average) if *pepper* is preceded by the word *salt* (compared to a control condition that precedes *pepper* by an unrelated word such as *loan*—Neely, VerWys, & Kahan, 1998). This is a large effect, given that a single word is easily read within about 200 ms from first sight (e.g., Rayner & Pollatsek, 1994). However, if *salt* is presented twice in succession, in the same task, just before *pepper* appears, the large facilitation effect vanishes (Balota & Paul, 1996; Neely et al., 1998). If this was merely an isolated oddball finding then it might be of little consequence, but all simple reading tasks reveal such complicated patterns of interaction among the factors that reading scientists study (see Bargh, 2006; Pickering & Ferreira, 2008, for reviews and discussions).

A slight variation in a laboratory task— the introduction of a new factor into an experimental paradigm—and the entire landscape of effects in reading performance can change (Van Orden & Kloos, 2005). Nonetheless, the default assumption in almost all reading research has been that the impact of a contributing factor to reading performance, whether the property of a text or of a reader, will be proportional to its magnitude. A less skilled reader should require proportionally more time and effort to read the same text, and a more difficult text should require proportionally more time and effort to read. But reading may comprise a heterarchy of overlapping and interacting capacities, such that different combinations may even compensate for deficiencies, insofar as reading speed or comprehension are concerned: This is well illustrated by an example given by Rayner and Pollatsek (1994) regarding the effects of speed reading, where it is generally found that very fast reader usually ‘skim’ though text, exhibiting
fewer fixations during reading, which in turn can have an impact on text comprehension. However, when for example a political leader reads through the first few pages of a daily newspaper, he or she might do so at the limit of speeded reading while maintaining a high level of content knowledge, simply because he or she was directly involved in most of the events portrayed on the first pages of the newspaper.

Accordingly, psycholinguists have come to understand that there exist complicated relationships between reader and text properties, but also even the most basic word descriptors, the lexical variables (such as word frequency), that putatively capture the cognitively salient features of words. Hence, caution is warranted when one confronts the scientific investigation of more extended text units, since one will face the above-mentioned complications all at once (e.g., Fisher & Shebilske, 1985; Kliegl, Olson, & Davidson, 1982).

Another reason why the investigation of individual words and sentences is so prominent is that they seem to encapsulate the essentials of written language: Words contain the basic meanings and lexical features of written language, while sentences supply the syntactical features, which can be sufficiently tabulated within a single sentence unit. Following this logic, the investigation of single (or a few) words and single (or a few) sentences contains the potential to uncover the basic features and rules of written language perception, which should be the principal basis for all reading performance (Carroll, 2003).

In an idealized lexicon, these constituents play a central role and are elaborately described to include meanings, spellings, pronunciations, and even the possible uses in sentence constructions (e.g., Bresnan, 2001). The constituents are elementary units
and their use in conjunction with grammatical rules depends on their unchanging character. They do not possess any interesting dynamics of themselves and their entries in the lexicon should not depend on any contextualization apart from what is already specified in their representation.

The pragmatic reason of handleability and the theoretic assumption that words or sentences are the basic constituents of reading behavior lead to the presumption that a characteristic scale of reading performance exists, which might be found on the word and/or sentences level, respectively. Therefore, only few studies have concerned themselves with reading on the text level or carry-over effects between sentences in a text (e.g., Just and Carpenter, 1980; Zwaan, Magliano, & Graesser, 1995), but even those have grounded their basic assumptions on the more abstract logic of concatenated effects (Michell, 1999). Concatenated effects are those that unfold in a time ordered fashion and share a common timescale. Thus, when a printed word is read aloud, the components and subcomponents of sensation, perception and reading are assumed to concatenate their effects like a row of time-ordered falling dominos, each affecting the next in its turn (Sternberg, 1969).

The abstract logic of concatenated effects justified ordinary contrasts between the average times of—for example—reading aloud different kinds of words. Yet, reading aloud is done by a brain and body in which ubiquitous feedback processes compose the nervous system, and in which simultaneous afferent and efferent processes control the eyes and mouth at the sensorimotor periphery of the reader. Moreover, feedback dynamics self-organize rapid and even instantaneous changes across the nervous system (Kelso, 1995; Wallot and Van Orden, 2011d), which puts the nervous system on
an entirely different range of timescales than the time course of repeated measurements of behavior (Van Orden, Kloos, & Wallot, 2011). If this organization dictates assumptions, then the measurement problem may be less one of timing dominos, and more one of characterizing loops and levels of interdependence in entangled cognitive phenomena (Bell, 1999).

The linguistic structure of a written text may also motivate loops and levels of interdependence in reading behavior. When words are read in succession, they influence each other in complex ways. As laid out in the above paragraphs, strict stimulus control is conventionally used to strip away the complication, and simple laboratory tasks and measurement procedures are chosen to minimize unwanted task effects. Nonetheless, even the simplest reading tasks embody complicated interactions in which the interacting factors may either amplify, reduce, or even eliminate each other’s effects.

If one want to understand reading performance by identifying a causal controller for each effect observed in reading performance, the picture of reading that is painted following these considerations looks very blurry, and indeed Rayner and Pollatsek (1994) have not been too optimistic about the attainment of a general theory of the reading process on the discourse and text level in the near future: “An eventual understanding of reading will clearly require, as an important component, an understanding of how discourse is represented and how this representation is constructed. However, we feel that our understanding of discourse representation has progressed little beyond that of the man or woman in the street, and we are not optimistic that there are any breakthroughs lurking around the corner” (p. 321).
However, it is worth to look at another discipline that faced similar problems regarding the complexity of the phenomena under investigation, and what has been learned from trying out new avenues: In the science of movement and motor control, the outlook was similarly dark. Here it was likewise assumed that each degree of freedom of the body must be controlled individually by a causal controller. For example, speech entails changes in about 70 muscles and speech activities such as conversations are also variable expressions of the body, entailing very many (vast) degrees of freedom. A source of controllers to control each degree of freedom would be quickly overwhelmed (Rączaszek-Leonardi & Kelso, 2008), a problem known as the degrees-of-freedom problem (Bernstein, 1967; Turvey, 2007).

The degrees-of-freedom problem has a counterpart in abstract descriptions of languages. Languages are abstract entities consisting of multiple parts, and combinations of parts, of various sizes and on different scales. As the number of described words, described words’ meanings, and the possible ordered combinations of words’ meanings grow, their number soon becomes indefinitely large, rivaling or trumping the number of neurons in a brain. The question is: What can be done?

Wallot and Van Orden (2011a) have argued that the scientific study of reading could follow the lead of movement science, and adopt a new view on the process of reading and cognitive activity in general (see also Van Orden, Holden, & Turvey, 2003; 2005) and change the focus from the isolation of causal factors in reading to the overall coordination and organization of reading activity. As will be laid out in the last section of this chapter, this will allow--in principle--to delve into the full-blown complexity of text reading without any detours, but will change the emphasis from the outcome of a
particular reading performance to the focus on the stability and interconnectivity of general reading performance.

The research presented in this dissertation aims at extending this perspective, starting with the work of Wallot and Van Orden (2011a, b). These studies have presented results from a self-paced reading task, during which participants advanced themselves through a text by revealing each new text unit via a key-press (see Figure 1). The data collected were inter-key-press intervals, which estimate the reading time of each text unit. In the following, the key-findings of these experiments will be summarized. Then it will be discussed how these findings speak to the problems in reading research discussed above. Finally, the chapter ends with an outlook on the studies presented in this thesis and how they expand on the work in Wallot and Van Orden (2011a, b).

Figure 1.1. Illustration of the self-paced reading task.
The factors investigated in Wallot and Van Orden (2011a, b) were differences in reading performance due to the length of the presented text unit (participants read a text either word-by-word, phrase-by-phrase, or sentence-by-sentence), differences in reading performance due to habitual differences in reading experience (participants were either undergraduate students in their first or seconds year, or Ph.D. candidates of English Literature), as well as differences in reading performance due specific differences in reading experience with a specific text (participants either read the story once or repeatedly). The latter two manipulations aimed at differences in reading fluency.

Regarding the size of the text unit presented, Wallot and Van Orden (2011a, b) found that word unit reading, even though it is the standard mode of presentation used in self-paced reading tasks, does not capture reading time in the common sense. That is, a 1-to-1 relation between word properties and the inter-key-press interval was not observed. Since all participants were literate readers, their reading ability outstripped their ability to reveal words using the response-key, which turned the reading task into a fast-paced tapping task in which the text merely influenced the stream of consecutive key-presses occasionally. The presentation of longer multi-word text units such as phrases and sentences showed a more reliable estimation of reading times (Wallot & Van Orden, 2011a), and especially the performance based on sentence unit presentation was sensitive to the differences between more and less experienced readers, as well as to the effects of repeated reading.

Regarding reading fluency, the sentence unit condition gave the following picture of the nature of the differences between participants with different levels of reading
experience and between reading the same text once or repeatedly: Overall, the reading times series of graduate students of English Literature showed greater entrainment to the reading task compared to the reading time series of undergraduates. Furthermore, the degree of entrainment of reading time series of graduate students either did not change appreciably with repeated reading or showed a slight decrease, while undergraduate students’ reading time series increased in entrainment with repeated reading. Wallot and Van Orden (2011b) interpreted this as indicating that graduate student readers had achieved a level of reading fluency that allowed them to read a--relatively easy--text as effortlessly at first sight as with repeated exposure. The additional knowledge through repeated reading did not appreciably improve their ease of reading, putatively because it was already at ceiling for this reader group reading a fairly easy piece of text. Undergraduate readers on the other hand showed a capacity of change in reading performance with repeated reading, which showed itself in an increase in entrainment.

The term entrainment refers to a hierarchical coupling process, where one system’s evolution follows the evolution of another leading system. An example of entrainment are circadian rhythms, where changes of an organisms activity and rest phases or body temperature follow the environmental change in light and darkness (e.g., Simpson & Galbraith, 1095). The term entrainment seemed to be a good choice as a leading concept to interpret complex reading performance for three reasons: First, entrainment focusses on the (co-)evolution of the text and the reader’s performance, highlighting the importance of dynamic properties instead of static stimulus-response relationships. Seconds, entrainment does not presume that the observed reading
behavior is caused by the text or its properties--it is as much a property of the text as well as a property of the capacities of the reader. Third, the notion of entrainment leaves open what it is in a text that a reader entrains to. On the positive side, this point eliminates the need to construe reading performance as unfolding on one specific scale, such as the number of words of a text. On the negative side, the term entrainment is in this aspect currently a mere replacement character for a more accurate description of what defines the psychological salience of a text then.

With the presented arguments in mind, the aim of this dissertation is to expand the research of Wallot and Van Orden (2011a, b) as follows: First, as we have seen, different text chunks are differently sensitive to manipulations in a self-paced text reading task. The sentence condition has proved to be the most promising. At the same time, sentence reading times are a relatively coarse-grained measure. Since word-by-word presentation effectively result a kind of tapping performance with skilled, literate readers and makes thus a poor task to measure reading, another measurement technique has to be sought to investigate the structure of text reading on faster time scales. At the same time, a different measure than button presses would also shed light on the generalizability of the findings observed from graduate and undergraduate readers under different task constraints. As pointed out, relatively slight changes in an experimental setup have the potential to change the measurement outcomes significantly, so the outcome observed from key-press responses need not necessarily generalize to other measures.

Hence, study 1 will use the experimental setup of Wallot and Van Orden (2011a, b) and collect eye-movements using the same procedure and stimulus text. Also,
another sample of undergraduate and graduate students who did the reading task as self-paced sentence-by-sentence reading was collected. This setup will allow for a partial replication of the results presented in Wallot and Van Orden (2011a, b) and complement them with a new measure—eye-movements—which will at the same time generate information about the structure of reading dynamics on faster time scales. Ideally, we would expect that differences in reading fluency result in like changes across multiple measures.

Study 2 aims at generalizing the findings of Wallot and VanOrden (2011a, b) to other stimulus materials, i.e., other texts, and will introduce a different kind of manipulation of text reading: Comparing reading performance between texts with different difficulty. Obviously, since the changes in reading dynamics have only been obtained using the same text material, it is warranted to question in how far the observed differences are due to idiosyncrasies of the stimulus material. Thus, two new texts have been selected for reading. Furthermore, the differences in text difficulty bring in a new manipulation of ease-of-reading: As described earlier, it is difficult—if not impossible—to dissect reading performance into isolable contributions by different text or task properties, because these properties are nonlinearly dependent on each other. Wallot and Van Orden (2011c) go even further: According to the authors, it is not only impossible to unambiguously parse up the variance due to task and language manipulations to obtain a stable picture of the performance of interest, but it is likewise impossible to cleave apart contributions of the task setup from contributions of the properties of the participant. Even though manipulations of so-called independent variables give the impression that factors—on the side of the task or the participant—can
be hand picked and varied individually, the nonlinear dependencies among task and participant properties have defied such an analytical approach (see also Van Orden, Pennington, & Stone, 2001).

In terms of reading performance this might mean that we might find the same differences observed due to reading skill (Wallot & Van Orden, 2011a, b) induced by differences in text difficulty for participants with a similar level of reading skill, if we can take the relevant nonlinearities into account. Of course, this remains an empirical question. Before we set out to attempt to answer these and other questions in the presented studies of text reading, I want to introduce the novel analysis techniques used to assess the nonlinear dynamic structures of reading performance, their rational and their working in Chapter 2. Chapters 3 and 4 will outline the experimental setup, results and conclusion of studies 1 and 2, respectively. Chapter 5 will bring these findings together, summarize what has been learned about the process of text reading, and discuss further research avenues.
CHAPTER 2:  
Methods and Motivation

In this chapter the nonlinear analysis methods that were used to analyze the reading data will be introduces. First, Mono- and Multifractal Analyses techniques will be described, and then Recurrence Quantification and Cross-Recurrence Analysis techniques will be described. The chapter ends with a summary of what the benefits of these analyses are when it comes to complex reading tasks.

Fractal analysis

Fractal analysis--as well all the other analysis techniques that are introduced in the following sections--requires time series data of at least several hundred data points, collected continuously, for reliable quantification of nonlinear structure. Fractal analyses provide two kinds of information to a scientist about the qualities of the data: The results of a fractal analysis serve as a test of whether it is justified to use a conventional analysis. And a fractal analysis estimates the fractal dimension ($FD$) of a data set, and fractal dimensions that differ reliably from $FD = 1.5$ (the fractal dimension of random white noise) may imply that no stable variance exists and sometimes that no stable average value exists (e.g., West & Deering, 1995). Conventional analyses grounded in the General Linear Model assume stable mean and variance and that data values are statistically independent, one from another, predicting data--or residual--series that vary
in a pattern equivalent to random white noise (Gilden, 1997, 2001; Gilden, Thornton, & Mallon, 1995).

The fractal dimension of white noise is a good example from which to get a feel for the concept of a fractal dimension more generally. The probability that one data point will be succeeded by a value greater than or less than the central tendency of the data is 50-50. Thus, about half the time a graph will include data points greater than the central tendency, and the other half of the time the data points will be less than the central tendency--so loosely speaking, the data graph fills up about half the space above the central tendency and half the space below the central tendency.

The fractal dimension of white noise, $FD = 1.5$, captures this idea if we imagine that white noise leaks halfway into the 2nd dimension of Euclidean space. The data graph of white noise is not simply a line--random noise is too irregular to be simply a line--but neither is it a 2-D object, it is ‘halfway’ in between. Fractal dimensions exist in between the canonical dimensions of Euclidean geometry. Euclidean dimensions are sufficient to capture the highly regular structures of Euclidean geometry, including 0-D points, 1-D lines, 2-D planes, and 3-D spheres or cubes. Fractal objects live between these canonical objects and the data series presented in Figure 2.1 is such an irregular fractal object, with $FD = 1.23$. The data graph is not a 1-D line, being highly irregular and aperiodic, and thus leaks somewhat from the 1st dimension into the 2nd dimension.
Figure 2.1. Top: The time series of one participant’s word reading times. Middle: The same time series, but with extreme outliers removed and normalized is subjected to spectral power density analysis (bottom) where sine and cosine waveforms are fitted to the data to estimate the power at different frequencies. The Power spectrum is plotted on log-log axis to investigate regions of scaling in the time series.
When a time-series of data has a fractal dimension reliably different from that of random white noise, then a conventional linear analysis may be unjustified and could thus be misleading. The fractal analysis in this case serves as a test of whether the necessary assumptions of conventional analyses are met. If this is not the case and the observed fluctuations differ reliably from white noise, then either the variance of the data series is nonstationary or both the mean and the variance are nonstationary, and nonstationary descriptive statistics cannot be trusted to be stable estimates of population parameters (Brown & Liebovich, 2010). In this case, fractal analysis can supply a second kind of information, which is the strength of long-range dependencies in the data, quantified by the value of the fractal dimension.

The fractal dimension of the time-series of word reading times in Figure 3, \( FD = 1.23 \), is very close to a fractal pattern called pink noise (or \( 1/f \) noise). Idealized pink noise has a fractal dimension of \( FD = 1.2 \), and thus leaks less into the 2nd dimension compared to white noise. Pink noise indicates nonstationary variance that will grow as more data are collected (e.g., Van Orden et al., 2011).

### Monofractal analysis

There are presently two kinds of fractal analyses applied productively to human event time data. Mono-fractal analyses estimate a single fractal dimension across an entire data set. They assume that a data series has a stationary fractal dimension. This assumption is not always justified, though. Human response time data are sometimes
monofractal, but more often the fractal dimension is nonstationary, changing reliably during the course of data collection (Ihlen & Vereijken, 2010).

The spectral analysis of each participant’s reading times yielded a log-log plot like that portrayed in Figure 2.1. The Y-axis in the plot tracks the magnitudes of variation as changes in reading times and the X-axis tracks how often changes of particular magnitudes occur. The relation between size of change and frequency of change is estimated by a least-square regression analysis, which quantifies the relation between size and frequency in the slope of the regression line. The end result estimates a scaling relation between size $S(f)$ and frequency $(f)$: $S(f) = 1/f^\alpha$.

The slope estimated in Figure 2.1 is negative, $-\alpha = -0.76$, which in turn gives us a scaling exponent of $\alpha = .76$. This value lies between the scaling exponents that correspond to idealized white noise (scaling exponent of $\alpha = 0$, slope = 0, fractal dimension $FD = 1.5$) and the idealized fractal pattern that we refer to as pink noise (scaling exponent of $\alpha = 1$, slope = -1, fractal dimension $FD = 1.2$).

Like other descriptive statistics, scaling exponents (and fractal dimensions) are most informative in contrasts that compare different outcomes, and laboratory performances can yield a wide range of scaling exponent values (Van Orden et al., 2011). Also, just as for other performance measures, the observed variation is due to tradeoffs among task demands, the integrity of the system, participant skills, experimental manipulations, and participants’ strategies (Kloos & Van Orden, 2010; Van Orden, 2010).
Multifractal analyses test whether a data set has a stationary fractal dimension or multiple fractal dimensions, varying across a multifractal spectrum. Multifractal analysis is an extension of mono-fractal analysis. It assesses specifically whether the variation in fractal dimension across a data series is substantially different from what would be expected of monofractal data. Thus a multifractal analysis tests whether variation changes in quality within the same data set. A positive outcome implies that a data series entails multiple different organizations of the system that performs a task.

Monofractal analyses assume that the pattern of variability across a time-ordered data set is sufficiently characterized by a single stationary $\alpha$ value. This is certainly not always - and perhaps it is not even often - the case. The more frequent outcome appears to be nonstationary $\alpha$ values that fluctuate within the course of a task performance (Ihlen & Vereijken, 2010). Multifractal analysis addresses the question of the quantity and the heterogeneity of $\alpha$ values that describe a data series (Brown & Liebovitch, 2010).

Mathematically, the tools of multifractal analyses are formal generalizations of the concept of mono-fractal analyses to higher moments of variability of response series (Ihlen & Vereijken, 2010; Mandelbrot, 1997). The monofractal $\alpha$ value can be loosely thought of as the central tendency of the multifractal case, its strongest scaling region. The spectrum of $\alpha$ values that result from a multifractal analysis can be seen as a dispersion of $\alpha$ values around that central tendency for different magnitudes of fluctuations in the data-series. A key statistic of interest is the width of the multifractal
spectrum, which defines the amplitude difference between the variability in the intermittent periods, where the values of the data series exhibits extreme fluctuations, and in the laminar periods of the observed time series, where fluctuations of values of the data series are minimal or absent. If the width is sufficiently small, the mono-fractal $\alpha$ value is a good enough estimate of the fractal properties of a data set. If the multifractal spectrum is sufficiently wide then it would be a mistake to ignore this essential character of the data, that its fractal properties are changing during the task performance and indicate that the underlying dynamics are changing in quality (Riley & Turvey, 2002).

Multifractal phenomena are unabashedly complex in the strict formal sense of complexity science, implying interaction-dominant dynamics among the component processes of the reading performance (Ihlen & Vereijken, 2010). The presence of multifractal dynamics in data may also imply that a studied system self-organizes its behavior to suite the requirements of the task (Kuznetsov & Wallot, 2011). Interaction-dominant dynamics allow the component processes of a system to change each other’s dynamics perpetually as they interact (Van Orden, et al., 2003). This creates interdependence among component processes, which allows context to constrain changes in all interacting processes, becoming constitutive of the self-organization (Van Orden, Kello, & Holden, 2010).
Recurrence Quantification Analysis

Recurrence Quantification Analysis (RQA) is another time series analysis technique that can quantify nonlinearities, such as the stability or chaoticity of a data series. Recurrence quantification analysis begins with phase-space reconstruction, which is essentially a projection of the time series into higher-dimensional space where the phase space portrait is inferred from the 1-D data series of reading times. The ordered reading times are plotted against themselves by a chosen constant lag—a delay parameter. For example, when a simple time-series is projected into 3 dimensions with delay 2, then the original time series is plotted against itself two more times, each time shifting the time series by two data points (the first datum is plotted against the 3rd and 5th in 3-dimensional space). This constructs a higher dimensional portrait of the system dynamics. The constructed space is used, in turn, to define structure in a 2-D recurrence matrix in which only equivalent reading times at equivalent locations in the phase space are marked. Webber and Zbilut (2005) is an accessible introduction to phase-space reconstruction in RQA; the mathematical rationale and proofs can be found in Takens (1981; see also Mañé, 1981; Ruelle & Takens, 1971).

The recurrence matrix is formed by following the order in which the data points were collected on both axes of a matrix, and then marking each point in the matrix at which equal reading times recur at the same location, along a shared trajectory, in the higher-dimension phase space. The recurrent reading times are thus located in the neighborhood of equivalent values along a shared trajectory of the space that was formed using lagged reading times.
Recurrence Quantification Analysis quantifies the geometry of the dynamics represented in the recurrence matrix. It yields variables that quantify the dynamics of a hypothetical system that could have produced the data series in question. The central concept of recurrence concerns how often a system repeats its behavior. Data that follow a regular pattern of variation repeat the same behavior often, producing a lot of recurrences and a simple regular geometry in the recurrence matrix. In more irregular data series, recurrences become less frequent and the pattern of recurrences itself can change across the recurrence matrix. Recurrence quantification analysis quantifies these aspects of the matrix geometry producing variables that can distinguish among different data sets.

The number of dimensions of the constructed space also estimates the number of active system components (i.e., active degrees of freedom) producing the variation in reading times. This works because continuously updating feedback creates coupling among the active degrees of freedom, sufficient to create interdependence among the components. Interdependence means that each change in each component is reflected simultaneously among all the components. Interdependence thus imbues the changes in each component with information about how the other components are changing. Because everything is connected to everything else in the enactment of a behavior, we can use the information latent in 1-D data to recover the higher-dimensional dynamics of the hypothetical dynamical system in question. Marwan (2011b) discusses contemporary issues in applying the analysis, including how many data points are necessary to conduct recurrence quantification analysis reliably. The range is wide: sometimes, several hundred data points suffice, sometimes tens of thousands of data
points are needed, depending on the complexity of the data series and the dynamics of the performance of interest.

Figure 2.2. From a one dimensional time-series to a three-dimensional reconstructed phase-space to a recurrence plot. The top row illustrates how a time-series plotted against itself at time $t_4$, $t_8$, and $t_{12}$ yields coordinates for a three-dimensional phase-space portrait of that time-series (i.e., $x = t_4$, $y = t_8$, $z = t_{12}$). The bottom row illustrates how a neighborhood (indicated by the red circle) in the phase-space gives rise to one or more points in the recurrence plot.
Figure 2.2 is a cartoon illustration of a 1-D data series converted into a 3-D phase-space portrait. The data plot in the upper-left quadrant of the figure has time (presentation order) on the X-axis and the data measurement on the Y-axis. The data plot is also marked with three dots to indicate the lag delay for re-graphing the data in a 3-D portrait. The delay parameter is the distance, on the X-axis, between the lagged points. The lagged points are used to re-plot the data series (against itself) in the higher-dimension phase space.

The delay parameter is selected so as to minimize the interaction between points of the measured time-series (maximize the information available across the lagged points). This parameter is commonly estimated using the first minimum of the autocorrelation function or else using the mutual average information function of the data series. With the delay parameter in the example set at four, the data series can be re-plotted. The Y-value of the left-most dot in Figure 2.2 defines the value on a new X-axis in 3-D, the Y-value of the middle point is the value on a new Y-axis, and the Y-value of right-most point is the value on a new Z-axis, resulting in the (x, y, z) point in 3-space pictured in the upper-right quadrant of the figure.

The dimension parameter defines the number of dimensions of the phase-space portrait that is (re-)constructed (the number of times the time-series is plotted against itself with the specified delay). The first minimum of a nearest-neighbor function is usually used to estimate the number of dimensions needed. The data points that are used in this procedure (delay times dimension) are lost for the quantifications. For example, if a data series of 100 points is plotted against itself with a lag of 10 (e.g., 1 vs. 10, 2 vs. 11, 3 vs. 12 (...) 90 vs. 99, 91 vs. 100), then the last 10 data points of the
time series cannot be plotted because there is no data point left in the corresponding
data series that has been shifted forward by 10 points. Hence, if dimension is a high
value and or delay is a large value, the remaining data points have to be sufficient in
number to reliably estimate quantities from the recurrence plot.

With all data re-plotted in 3-space (the spiral shape of the upper-right quadrant
in Figure 2.2), a neighborhood criterion is chosen--the radius parameter--portrayed by a
circle in the lower-left quadrant. The radius parameter sets a spherical threshold within
which points in the reconstructed phase-space portrait are counted as equivalent
recurrent points. The radius is chosen conservatively to define the neighborhood
equivalence around each point in the phase-space, which in turn decides which of the
nearby points should be counted as intersecting trajectories (given a little noise).
Generally, the radius parameter is chosen so that the recurrence rates do not exceed
5% (Webber & Zbilut, 2005). If the radius parameter gets too big, then points in the
phase-space will to be counted as being identical, when in fact they are not.

Five recurrent points are illustrated in the recurrence matrix in the bottom-right
quadrant of Figure 2.2. The recurrence plot is fully fleshed out by marking each
intersecting point of recurrent values with a dot. With all these dots in place, the larger
structures of the recurrence plot become visible as lines of dots, ‘empty spaces,’ and
larger geometrical structures. The visible structure provides the basis with which to
quantify the nonlinear variables of the recurrence quantification analysis.

Recurrence Quantification Analysis yields multiple quantitative outcome
variables, and more are currently under development. Not all outcome variables are
always applicable or meaningful for every data set (see Marwan, 2011b). Hence, one
must pick measures that make sense in the context of the data at hand and the research questions to be answered. Some measured values such as recurrence rate (\%RECurrence) or the amount of deterministic structure in the data (\%DETerminism) are meaningful in most contexts, and for the forthcoming analyses I examined \%RECurrence, \%DETerminism, \%LAM, and ENTRopy of diagonal lines.

With the radius parameter kept constant, a higher \%RECurrence indicates that the behavior is less spread out in the phase space, perhaps less noisy or confined to fewer different states. \%DETerminism equals the number of diagonally adjacent points in the recurrence plot divided by the total number of recurrent points, which estimates the degree of order in the data. High \%LAMinarity indicates that there are only few overall changes in the dynamics of the responses over time, that the performance is confined to only a few different states. Low \%LAMinarity on the other hand indicates that changes over time are erratic. Finally, ENTRopy of diagonal lines captures the diversity of the lengths of trajectories in phase-space, and is thus a measure of the complexity of the dynamics, with higher ENTRopy indicating lower complexity of transitions and vise versa.

Cross-Recurrence Quantification Analysis

Cross-Recurrence Analysis (CRQ) extends recurrence analysis in much the same way that a correlation between two random variables extends auto-correlation of a single random variable. In the latter case, we evaluate a data series against itself, and
in the former case against another data set. In this light, Cross-Recurrence Analysis can be thought of as a kind of nonlinear correlation analysis developed to test whether, and to what extent, two systems produce dynamics in common (Shockley, Butwill, Zbilut, & Webber, 2002). If they do, then common sources of constraints are implicated, where a change in constraints is associated with a change in the degrees of freedom of trajectories in the state space. More tightly constrained dynamics have fewer degrees of freedom and less tightly constrained dynamics have more degrees of freedom.

The procedure for conducting the Cross-Recurrence Analysis analysis is nearly the same as that for conducting simple Recurrence Quantification Analysis, the primary difference being that two data sets are contrasted instead of contrasting one data set with itself. That being the case, it can be crucial to normalize both data sets prior to analysis (for example by computing z-scores) so that they share a common scale. Shockley (2005) gives a good introduction to cross-recurrence quantification analysis (and see also Marwan, 2011b). In this dissertation, Cross-Recurrence Analysis is used to compare the shared dynamical structure among data sets collected in the same condition. The idea behind this analysis is that behavior that is enacted under similar constraints may possibly entrain to the present task in a similar way. Such commonalities may yield parallel dynamics that can be captured in Cross-Recurrence Analysis.
What Can and Do Nonlinear Analysis Techniques Offer?

After having described the analyses methods in more detail, what picture of the performance under investigation do they paint, and what can we hope to achieve with the use of these methods?

These time-series analysis techniques change the focus from the information contained within a change in central tendency, which is usually the unit of interest when quantifying component effects, to the information shared among many data points, across a whole task performance. Also, these methods allow a new emphasis in the interpretations of differences in stability, instability, and interconnectivity among system components. This may mean that cognitive activity will be re-interpreted, since new terminology must be adopted in order to make use of, and properly conceptualize, the new information extracted from the data.

The promise of these efforts is a richer and more detailed picture of cognitive performance, which has eluded the component dominant analysis of reading performance so far. Hence, the following studies of text reading do not only aim at deepening the scientific understanding of the reading process, but also at showcasing how to utilize these methods to investigate cognitive performance in general.
CHAPTER 3:

Study 1 - Repeated Text Reading: Key-Response Times and Eye-Movements

As has been fleshed out in the introductory Chapter, the first study incorporates eye-movement measures into the investigation of the dynamics of text reading. Results are related to previous investigations (i.e., Wallot & Van Orden, 2011a, b), emphasizing the studies of Wallot and Van Orden (2011b) that showed that the dynamics across sentence reading times in a self-paced reading task were most predictive of differences in reading fluency.

In particular, the sentence reading times of undergraduate students—the participant group assumed to have less reading experience—showed stronger entrainment to the text with repeated reading, while graduate students of English Literature—the participant group assumed to have the most reading experience—showed either no change in reading performance over the subsequent reading or a trend toward more idiosyncratic performance.

It is interesting to note that sentence unit presentation was more revealing of differences between reader groups than word unit presentation, the much more commonly used mode in the self-paced reading tasks of prior research. Wallot and Van Orden (2011a) attribute the lack of sensitivity of word unit presentation to the fact that competent readers can read words faster than they can reveal each word with a button press. Hence, performance observed in the word unit presentation mode can be seen as being a mesh of reading times and, mostly, a fast-paced tapping task. Nevertheless, the presence of effects of reading experience in sentence-unit reading was somewhat
unexpected, since words are usually assumed to be the more fundamental unit in reading (Rayner & Pollatsek, 1994), while sentences are already a conglomerate of words, perhaps subject to sequential priming effects and higher-order syntactic influence, even though the latter point has come under recent scrutiny (Frank & Bod, 2011).

Since self-paced word unit presentation elicits ‘tapping performance’, a different measure had to be sought to investigate reading performance on a more fine-grained scale. The measurement of eye-movements is the next promising candidate. Eye-movements have a comparably long history within reading research (Rayner, 1998), and the investigation of eye-movements in reading has revealed details about the reading process that would be difficult, if not impossible, to measure with conventional response-key-based designs. For example, that about 10% to 15% of the saccades during reading are regression, returning to previously read text (e.g., Murray & Kennedy, 1988) or that words are frequently skipped during reading, but can--at the same time--be partially perceived parafoveally at the edges of fixations (e.g., Inhoff & Radach, 1998). Furthermore, Rayner and Duffy (1986) have presented evidence that the variability in eye-movement measures is indicative of cognitive activity during reading, a fact that the experiments in this dissertation want to capitalize and expand on.

Fortunately, as Rayner and Pollatsek (1994) pointed out, there is quite a bit of convergence observed in the effects of word- and text-related manipulations across self-paced reading studies and eye-movement studies of reading. Hence, eye-movements are not only a promising measure of reading performance by themselves, but there is reason to suspect that somewhat similar patterns of reading performance
might be observed across measures of eye-movements and those of key-presses in
self-paced reading.

These considerations motivate the first reading study: The text material from
Wallot and Van Orden (2011a) were used to replicate their findings regarding reading
fluency differences between groups of readers with different reading experience, as well
as between the first and second reading of a text in the self-paced sentence unit
presentation mode. In addition to that, eye-movement recording was employed to
validate the effects of repeated reading.

In the following sections the experimental setup will be introduced, together with
the procedure and materials used. After that, results of conventional and nonlinear
analyses of key-presses and eye-movements will be presented and discussed. Finally,
findings across these two measures will be summarized and discussed, also with
respect to the previous studies of Wallot and Van Orden (2011a, b).

Method

Design and Participants

The study follows a semi-experimental design with two factors: Participants were
either graduate students or undergraduate students and each participant read the story
either one or two times. Hence, the design conformed to a 2 (reading experience:
graduate vs. undergraduate/between) by 2 (number of readings: first vs. second/within)
factorial structure. The performance of participants was either recorded via key-presses
on a keyboard or via eye-tracker.
All participants were students of the University of Cincinnati. Overall, 42 students participated in the study. 18 were graduate students of English Literature, Philosophy, Psychology, and Sociology (average age 26.47 yrs., ranging from 22 to 41; 4 males). The remaining 24 students were undergraduate students of Psychology (average age 19.57 yrs., ranging from 18 to 24; 6 males). All were native speakers of English and had normal or corrected-to-normal vision. Of the 42 students, 8 had to be dropped, because the eye-tracking equipment did not pick out their eye-movements reliably. Sixteen of the remaining students (8 of them graduate students) took part in the full-page presentation condition where participants’ eye-movements during reading were measured. The other 18 students (7 of them graduate students) took part in the sentence unit presentation condition, where participants advanced themselves through the text by pressing a space bar to reveal every new sentence. All participants were compensated in cash for their participation.

Stimuli and Apparatus

The text used was the same as in Wallot and Van Orden (2011a, b), a short story entitled ‘The Arelis Complex’ by Louis P. DeGrado (2003). This story describes the fictive intergalactic politics of the Arelians, who were nearly wiped out in a previous conflict with another civilization. Star ship officer Drakh Norh, charged with the safety of the Arelian home planet, oversees the interception of an unmanned space vessel. The vessel is traced back to a ‘blue planet’ whose inhabitants had been deemed incapable of such technology. Officer Drakh concludes that the inhabitants of the blue planet may
eventually develop into a threat and leads a war fleet to destroy its population. The attempt fails, however, leaving the Arelians facing an uncertain future.

The Arelis Complex consists of 1,042 sentences. The average sentence length is 13.36 words ($SD = 7.89$), lying well in the range of standard English prose (Fengxiang, 2007; Sigurd, Eeg-Olofsson, & van de Weijer, 2004). The Flesch-Kincaid index of readability assigned the text a score of 6.0, indicating that it’s difficulty is appropriate for early 6th grade readers (even though readability formulas are not without problems (Chall, Bissex, Conard, & Sharples, 1996; Fry, 2002), the text as such should therefore make an easy reading for college students at any stage).

Text units were presented in Courier New font (14 pt.) on a standard 13-inch computer monitor. The program that controlled stimulus presentation was a MatLab© PsychophysicsToolbox (Brainard, 1997; Pelli, 1997) script, run on a standard PC.

In the self-paced reading condition, participants advanced themselves through the text by pressing the space bar of a computer keyboard, revealing every new sentence in turn. The data collected were the intervals between two consecutive space bar presses. These are estimators of the time it took a participant to read a sentence.

In the eye-movement recordings condition, participants saw a whole page of text, displaying on average at about 140 words. Participants used the space bar to flip pages, revealing a new page full of text with every button press. During reading, the eye-movements of participants were monitored by a AppliedScienceLaboratory D6© remote eye-tracking system with a temporal resolution of 60Hz and a maximal spatial resolution of 0.5° visual angle.
Two versions of a comprehension questionnaire were used to ensure a minimum of reading motivation on the side of participants. Each version consisted of 10 multiple-choice questions, targeting passages throughout the whole text. Furthermore, participants were asked to name the main characters and give a brief summary of the story. The purpose of the questionnaires was to ensure participants’ motivation to perform the reading task properly and the gathered data were not further analyzed (see also Zwaan et al., 1995).

Procedure

Participants were seated in front of the computer monitor that was used to display the text. After obtaining consent, participants were told that their task would be to read a short story, and that the story would be displayed on a computer monitor. Depending on condition, they were either instructed to press the space bar to reveal the story sentence-by-sentence or page-by-page. Participants were asked to only reveal the next text unit after they had finished reading the one that had been previously displayed. Also, participants in the full-page condition were told that their eye-movements would be measured during reading.

Before starting the reading task, participants were informed that they would be asked to complete a comprehension questionnaire related to the contents of the story after reading. It was also pointed out to them that there was no need to try to memorize specific parts of the story and that merely reading it would be sufficient in order to successfully answer all comprehension questions.
The eye-tracking system was calibrated to each participant by showing him or her a display of 17 dots, one at a time, at which participants what to look briefly in order to provide the eye-tracking system with reference points to calculate gaze position.

After reading the story, participants were given one of the versions of the comprehension questionnaire. With two exceptions, all participants returned 7 days later at the same time of day to read the story for a second time. Participants were told that they would re-read the same text again and the identical instructions were given. After reading the story once more, participants were given the other version of the comprehension questionnaire, debriefed and thanked.

Data preparation

Averages and standard deviations of sentence reading times, as indicated by the interval between two consecutive key presses, as well as averages and standard deviations of fixations and saccades were computed on the raw data. Also, raw data were used to conduct Multifractal Analysis and Recurrence Quantification Analysis. However, all data sets were trimmed to the first 4000 fixations and saccades for Cross-Recurrence Quantification Analysis, since the technique requires all paired data sets to be the same length. The Commandline Recurrence Plots software was used to conduct the recurrence analyses (Marwan, 2011a).

The Multifractal Detrended Fluctuation Analysis technique was used to estimate multifractality in the data series (Kantelhardt et al., 2002). To estimate fractal properties of a time series, Detrended Fluctuation Analysis breaks a data series into non-overlapping bins of a certain (small) bin size. Linear (or other) trends are removed from
each bin and the variance of the resulting residuals is calculated. Then, the size of the bins is increased and the procedure is repeated. This allows an estimation of the change in variance of a data series over different scales (bin sizes) to estimate the scaling properties of the fluctuations in that time series (Peng, Havlin, Stanley, & Goldberger, 1995). Multifractal Detrended Fluctuation Analysis is an expansion of this technique by which larger and smaller fluctuations in the time series are investigated in the same manner. If small, medium, and large fluctuations in a time series exhibit the same scaling properties, then the time series exhibits simple mono-fractal behavior. If the different fluctuations scale differently with bin size, then the data series is multifractal, and the degree of multifractality can be estimated by the range of the different scaling exponents observed, the multifractal spectrum. Since the presented data are considered ‘short’ data series with regards to the demands of this analysis technique (Kantelhardt et al., 2002), the $q$-parameter, which dampens or amplifies fluctuations in the time series of interest is restricted to the range of 0.1 to 3.0 (Ihlen & Vereijken, 2010) to avoid spurious results.

To conduct Monofractal Analysis, the recommendations of Holden (2005) were followed to prepare the data: All key-press intervals greater than 20,000 ms and smaller than 100 ms were removed and linear and quadratic trends were removed before the data series were subjected to Monofractal Analysis.

To separate saccades and fixations, an acceleration based criterion was used (Araujo, Kowler, & Pavel, 2001) where increases in acceleration of the eye-movement record indicate the onset of a saccade, while decreases in acceleration indicate the end of a saccade and the onset of a fixation. The the algorithm that implemented the
acceleration-criterion based separation procedure was taken from Hidalgo-Sotelo (i.e., Rich, Kunar, Van Wert, Hidalgo-Sotelo, Horowitz, & Wolfe, 2008) and operated in the following way: The time-series of eye-movement positions was divided into overlapping windows of adjacent data points with a window size of 4 (equalling 68 ms) and a step-size of 1 (equalling 17 ms). The acceleration threshold was set to a change in velocity of 2° of visual angle per second between two consecutive intervals (Abrams, Meyer, & Kornblum, 1989). The onset of a saccades was marked by an exceedance of this threshold. Fixations were defined as the periods between saccades with a minimum duration criterion of 51 ms. Successive saccades that occurred within 35 ms of each other were considered to be a continuous saccade. Also, for the purpose of the non-linear time series analyses, forward and backward saccades were not distinguished. This was done in order to preserved the maximal information in the saccadic dynamics.

For the eye-movement data, all fixations longer than 1,000 ms and all saccades slower than 250 ms were removed. The high threshold was adopted to account for the wide range of durations observed in fixations and saccades within and between individuals (Rayner, 1983), but to ensure at the same time that spurious fixations (e.g., larger than 15 seconds) were removed before analysis. All data points greater or smaller than 3 standard deviation from the mean were removed as well. Finally, trends up to quadratic were removed from the time series. This procedure ensures that analyses that estimate monofractal structure are not biased by singular data points or simple trends. Detrending also reduces the likelihood of a false rejection of conventional analysis, and if the pattern of reading times is truly fractal then the removal of outliers
and trends should not greatly impact the results (Caccia, Percival, Cannon, Raymond, & Bassingthwaite, 1997).

After preparing the data as described and calculating the key-statistics for each time-series, the resulting data set was subjected to a mixed repeated measures ANOVA with the between-measures factor reading experience (levels: graduate vs. undergraduate) and the within-measures factor number of readings (levels: one vs. two).

Results of Key-Press Intervals

Averages and Standard Deviations

As shown in Figure 3.1, average sentence reading times decreased with repeated readings for both, graduate and undergraduate readers ($F(1, 16) = 12.26, p < .01$). Furthermore, the average standard deviations for sentence reading times decreased with repeated reading as well ($F(1, 16) = 10.78, p < .01$). No effects of reading experience or interactions between reading experience and repeated reading were evident (all $F < 1.53$).
The decrease in average reading times indicates that there is a gain through re-reading the text, making it possible to read it in less time. The decrease in standard deviation indicates tighter dispersion of reading times around the central tendency.

![Figure 3.1](image)

Figure 3.1. Average sentence reading times decrease with repeated reading (left). Average standard deviations of sentence reading times decreases with repeated reading (right). Averages are bracketed by the standard errors of the means.

**Mono- and Multifractal Analyses**

The results of mono-fractal analysis were obtained using Power-Spectral Density Analysis (PSD), resulting in an alpha exponent that captures the mono-fractal properties of the time series of sentence reading times. As pictured in Figure 3.2, there is no reliable change in alpha exponents with repeated reading and no reliable difference between graduate and undergraduate readers, nor an interaction between the two factors (all $F < 0.53$). This indicates that simple monofractal long-range correlations remain stable over differences in reading fluency in sentence-unit reading. And the
overall magnitude of alpha values indicates that long-range dependencies are generally weak and close to white noise.

The multifractal spectrum width of sentence reading times did not reliably differ between graduate and undergraduate readers or between the 1st and 2nd reading (all $F < 1.60$).

![Figure 3.2](image.png)

Figure 3.2. Neither alpha exponents (left) nor multifractal spectrum widths (right) differ reliably as a function of reader group or repeated reading. Averages are bracketed by the standard errors of the means.

To test whether the average scaling exponent alpha, or the observed average width of the multifractal spectrum of one condition is reliably different from zero, a so-called surrogate data test can be performed. To conduct a surrogate test, each data set is paired with a shuffled copy of itself. The shuffled copy is obtained by randomizing the data points of the original time-series. This new, randomized time-series, the surrogate, retains the same central tendency and distribution of the original, but all temporal information is destroyed by the process of randomization. Monofractal and Multifractal Analyses can now be performed on the surrogate data, and a paired sample t-test can be applied to test the scaling exponent values of the original time-series against the
scaling exponent values of their surrogates for each condition. If the t-test indicates a reliable difference between the surrogates and the original time-series, then it can be assumed that the tested time-series data do not merely consist of random fluctuations, but contain long-range dependencies (see also Holden, 2005; Kantelhardt et al., 2002).

Surrogate analysis of monofractal alpha exponents revealed that all conditions produced alpha exponents reliably different from their the alpha exponents of their surrogates (all $t(7) > 2.93$, all $p < .05$). Surrogate analysis of the multifractal spectra of key-press intervals revealed likewise that all conditions yielded multifractal spectra widths that were reliably different from their surrogates (all $t(7) > 3.62$, all $p < .01$).

Recurrence Quantification Analysis

Before recurrence measures were computed, appropriate delay and dimension parameters for the reading time series had to be set. A delay of 1 time step (trial) was chosen for all the recurrence analyses due to the discrete nature of the measurements. The dimension parameter was estimated by the first local minimum of the false-nearest-neighbor function for each data set. The resulting estimators of dimensionality were averaged, resulting in an average embedding dimension of 4. This same (average) dimension parameter was used for all data sets. The radius parameter was set to yield an average of 5% recurrence for all data sets.

As outlined in Chapter 2, the recurrence quantifiers %RECurrence, %DETerminism, %LAMinarity, and the ENTRopy of diagonal line lengths were used to characterize the dynamics of reading time. This was done for three reasons: First, it allows for a comparison of the results with those presented by Wallot and Van Orden
(2011a, b) that used the first three of these variables as well. Second, %RECurrence, %DETerminism, %LAMinarity, and ENTRopy cover the range of currently available recurrence quantifiers, indexing the dispersion of states in phase-space (%RECurrence), the regularities in the transitions between states in phase-space (%DETerminism), the change (or chaoticity) of transitions between states in phase-space over time (%LAMinarity) and the overall complexity of phase-space trajectories (ENTRopy). Third, all other recurrence variables yielded either redundant results to those presented or null-results, thus offering no additional insights into the reading dynamics.

Figure 3.3. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) differ reliably as a function of reader group or repeated reading. Averages are bracketed by the standard errors of the means.
As can be seen in Figure 3.3, none of the four recurrence quantifiers of sentence reading times differs reliably with repeated reading or reading experience (all \( F < 2.50 \)).

As was to be expected from Wallot and Van Orden (2011b), recurrence analysis did not yield any insights into the reading dynamics of sentence reading times. In the general discussion, possible reasons for the failure of recurrence analysis to yield reliable differences between repeated reading and the groups of graduate and undergraduate readers will be addressed. However, there is another way in which the basic idea of recurrence analysis can be used to shed light on the different reading conditions, and this is by subjecting the data to cross-recurrence quantification analysis.

**Cross-Recurrence Quantification Analysis**

Cross-Recurrence Analysis can be used to look at commonalities among readers in a group that are due to the common accommodation of task demands, including how the text-unit conditions affect readers similarly or differently. By employing Cross-Recurrence Analysis, the structure that different readers have in common is highlighted in contrast to the whole structure contained within their performance which would include each reader’s idiosyncrasies.

Eleven undergraduate and 7 graduate participants resulted in \( 11!(2!(11-2)!)=55 \) possible pairings for undergraduate readers, and \( 7!(2!(7-2)!)=21 \) possible pairings for graduate readers. The same parameters (delay of 1, dimension of 4, 5% (cross-) recurrence) as in simple recurrence analysis were used to estimate the recurrence quantifiers. Figure 3.4 gives an overview of the results.
All recurrence quantifiers yielded a reliable effect of reading experience:

%RECurrence is reliably higher for graduate readers' performance compared to that of undergraduate readers ($F(1, 64) = 26.54, p < .001$). The same is true for %DETerminism ($F(1, 64) = 22.70, p < .001$), %LAMinarity ($F(1, 64) = 37.56, p < .001$), and ENTRopy ($F(1, 64) = 28.22, p < .001$). No other effects were apparent (all $F < 2.60$).

These findings indicate that the performances of graduate readers share more structure in their time course compared to undergraduate readers, and this irrespectively of the numbers of reading.
Discussion of Key-Press Intervals

Sentence reading times measured by key-presses dropped with repeated reading of the text, just as the standard deviations of key-press intervals. The facilitative effect of repeated reading goes along with faster, and overall less dispersed, reading times. Neither Monofractal Analysis nor Multifractal Analysis were sensitive to differences between the factors. Cross-Recurrence Analysis highlighted differences between reader groups, with recurrence measures being generally higher for graduate readers, indicating that their performance showed greater commonalities in their temporal evolution as indicated by higher cross-recurrence measures, compared to that of undergraduate readers.

How do the measures of key-presses relate to the findings of Wallot and Van Orden (2011a, b)? As far as the averages and standard deviations are regarded, the results of study 1 are only a partial replication of the original findings: While Wallot and Van Orden (2011b) also reported a decrease in reading times with repeated reading, they generally observed faster reading times for graduate compared to undergraduate student readers, which was not replicated in the current data. Why are we not seeing faster reading times for graduate students in this study? There are at least two viable explanations: First, Wallot and Van Orden (2011b) recruited graduate students from English Literature, while the graduate students who participated in this study came from a variety of fields, namely Psychology, Sociology, Philosophy, and English Literature. It might be that the groups of students from Psychology, Sociology and Philosophy are on average simply not as experienced readers as those studying English Literature. This
might have weakened the manipulation of reading fluency. Also, since participants were not urged to read the text fast, but at a comfortable speed, one participant, a graduate student from Sociology, took over 2 hours to read the story. This considerably increased the average reading time and standard error of the group of graduate readers. However, when the participant’s data was excluded from analysis, the effects remained the same.

The nonlinear effects were also only partially replicated. Especially, a crucial interaction effect reported for the change in nonlinear measures was not obtained. Wallot and Van Orden (2011b) found that measures of structure and stability of the performance of graduate readers obtained Cross-Recurrence Analysis remained relatively stable or decreased with repeated reading, while the same measure increased for undergraduate readers’ performance with repeated reading. This was interpreted as showing the undergraduates’ greater capacity for gains in reading fluency though repeated reading. In contrast, the current study reveals only a main effect of reader experience, showing more common regularities in graduate students’ key-presses compared to undergraduate students’ key-presses.

One explanation for this difference might be that in Wallot and Van Orden (2011a, b), participants did an additional reading of the text within 1-3 days after their first reading, whereas in the current study, participants come back after 7 days to re-read the story. Reading fluency develops across a time scale of years, from elementary school to college and beyond, while the effect of repeated reading may decay with relatively short half-life and is specific to the particular corpora or texts that have been read (Samuels, 1979). Compared to the more slowly developing reading fluency, the method of repeated readings has a transient effect on performance. Therefore, 7 instead of 1-3
days between the first and second reading could have greatly diminished the effect. This hypothesis gains credibility when we compared the differences in means and the overall magnitudes of the cross-recurrence measures between this study and Wallot and Van Orden’s (2011b) results: Even though a reliable decrease in reading times with repeated reading was found in the present study, the gain for graduate and undergraduate students is only about 450 ms per sentence, while graduate students gained about 650 ms per sentence and undergraduate gained about 1,600 ms per sentence in the Wallot and Van Orden (2011b) study.

Furthermore, when we look at the first reading in Wallot and Van Orden (2011b), graduate students’ performance possesses more commonalities compared to that of undergraduates--the same difference we observe in the present data throughout the first and second reading. Thus, it seems that the performance of graduate and undergraduate students during a second reading of the same text after 7 days is much more similar to the performance during their first reading, in terms of the overall speed as well as the structure of the performance. The gains in reading fluency through repeated reading of undergraduate students observed by Wallot and Van Orden (2011a, b), that were archived within 1-3 days, have been mostly lost again at 7 days.

**Results of Fixations and Saccades**

**Averages and Standard Deviations**

The average of fixations and saccades, as well as the average standard deviations of fixations and saccades are displayed in Figure 3.5. Both, the average
duration of fixations ($F(1, 14) = 7.86, p < .05$) and the average standard deviation of the duration of fixations ($F(1, 14) = 6.67, p < .05$) decrease reliably with repeated reading. No effects of reading experience and no interaction between the factors was found (all $F < 0.75$).

Figure 3.5. Average duration of fixations decreases reliably with repeated reading (upper-left). Average standard deviation of fixations decreases reliably with repeated reading (upper-right). Average duration of saccades generally increased with repeated reading, but the increased was smaller for graduate readers compared to undergraduate readers (lower-left). Average standard deviation of saccades did not differ reliably as a function of reader group or repeated reading (lower-right). Averages are bracketed by the standard errors of the means.

The average duration of saccades ($F(1, 14) = 6.05, p < .05$) increased reliably with repeated reading, and the average standard deviation of the duration of saccades showed a marginally reliable decrease with repeated reading ($F(1, 14) = 3.52, p = .082$).
However, this main effect on saccade durations was qualified by reliable interactions between the factors ‘repeated reading’ and ‘reading experience’: With repeated reading, the average duration of saccades \(F(1, 14) = 5.02, p < .05\) increased disproportionally for undergraduate readers; and a similar statistical tendency was observed for the standard deviation of saccade durations \(F(1, 14) = 4.01, p = .065\). There were no main effects of reading experience on saccade duration and dispersion (all \(F < 1.58\)). The results indicate that participants fixated words more quickly during the second reading. Also, the transitions between fixations, as indicated by saccade-duration, increased with repeated reading, and this increase was reliably pronounced in the undergraduate readers’ eye-movements.

![Fixations vs Saccades - Average Number](image)

Figure 3.6. The number of fixations dropped reliably with repeated reading (left). The number of saccades dropped reliably with repeated reading (right). Averages are bracketed by the standard errors of the means.

An analysis of the number of fixations and saccades reveals an effect of repeated reading as well (see Figure 3.6): For the second reading of the text, graduate and undergraduate readers exhibit fewer fixations \(F(1, 14) = 5.66, p < .05\) and saccades \(F(1, 14) = 5.66, p < .05\) compared to the first reading. This suggests that the
increase in saccade duration with repeated reading translates into longer leaps between fixations. There were no other effects ($F < 2.81$).

**Mono- and Multifractal Analyses**

The results of Mono-Fractal Analysis for fixations and saccades were calculated the same way as for key-presses. As can be seen in Figure 3.7, there is no reliable change in alpha exponents with repeated reading and no reliable difference between graduate and undergraduate readers, and also no interaction between the two factors (all $F < 2.25$). Just as with key-press performance, this indicates that simple monofractal long-range correlations remain stable over differences in reading fluency when measures through eye-movements, and the overall magnitude of alpha values indicates an absence of long-range dependencies, since the observed fluctuations are generally white noise.

The was no effect of repeated reading or reading experience on the multifractal spectrum width of fixations (all $F < 1.24$), but there was a marginally reliable interaction effect between these two factors ($F(1, 14) = 3.63, p = .077$), indicating a potential decrease in multifractality for graduate readers with repeated reading, while multifractality in undergraduate readers’ performance showed a potential increase with repeated reading. The multifractal spectrum width of saccades decreased reliably with repeated reading ($F(1, 14) = 4.70, p < .05$). There were no effects of reading experience and no interactions between the factors reading experience and repeated reading (all $F < 0.01$). Increased multifractality points to an increase in the heterogeneity of variance in reading times. The observed decrease in multifractality of saccades with
repeated reading indicates that the transitions between fixations showed less heterogenous variance when the text was read for the second time.

![Graphs](image)

**Figure 3.7.** Neither alpha exponents of fixations (upper-left) nor multifractal spectrum widths of fixations (upper-right) differ reliably as a function of reader group or repeated reading. Alpha exponents of saccades do not differ as a function of reader group or repeated reading (lower-left). Multifractal spectrum widths of saccades decrease reliably with repeated reading (lower-right). Averages are bracketed by the standard errors of the means.

Surrogate analysis of the monofractal alpha exponents of fixations revealed that all conditions yielded multifractal spectra widths that were reliably different from their surrogates (all $t(7) > 2.38$, all $p < .05$) with the exception of graduate readers saccades during the second reading, whose multifractal spectra widths were not reliably different from the multifractal spectra widths of their surrogates ($t < 1.34$). Surrogate analysis of
the multifractal spectra of fixations revealed that all conditions yielded multifractal spectra widths that were reliably different from their surrogates (all \( t(7) > 2.55, \) all \( p < .05 \)).

Surrogate analysis of the alpha exponents of saccades were generally not different from their surrogates (all \( t < 1.70 \)). Surrogate analysis of the multifractal spectra of saccades revealed that all conditions yielded multifractal spectra widths that were reliably different from their surrogates (all \( t(7) > 3.08, \) all \( p < .05 \)) with the exception of graduate readers saccades during the second reading, whose multifractal spectra widths were not reliably different from the multifractal spectra widths of their surrogates (\( t < 1.40 \)).

Recurrence Quantification Analysis

Recurrence Quantification Analysis was conducted on the fixations and saccades in the exact same way as described for the key-press intervals. The delay parameter was set to 1, the average embedding dimension of the data series was again obtained from the false-nearest-neighbor function. It was 5 for fixations and 6 for saccades. The radius parameter was set to yield an average of 5% recurrence for all data sets.
Figure 3.8. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) of fixations differ reliably as a function of reader group or repeated reading. Averages are bracketed by the standard errors of the means.

Figure 3.8 displays the recurrence quantifiers obtained from the time series of fixations. There are no statistical differences between any of the measures for repeated reading and reading experience, and there was no interaction between the factors (all $F < 1.20$).

Similarly, there were no reliable differences for repeated reading and reading experience, and no interactions between the factors for recurrence quantifiers of saccades ($F < 1.49$), see Figure 3.9.
Figure 3.9. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) of saccades differ reliably as a function of reader group or repeated reading. Averages are bracketed by the standard errors of the means.

Cross-Recurrence Quantification Analysis

Cross-Recurrence Analysis was also used to look at commonalities among readers in a group. As with the key-press intervals, the same parameters were used for cross-recurrence quantification analysis that estimated for recurrence analysis. Since there were 8 readers in each condition, there will be 8!/(2!*(8−2)!)=28 paired data sets in each bin.

As can be seen in Figure 3.10, %RECurrence (F(1, 54) = 3.89, p = .054) and %DETerminism (F(1, 54) = 5.42, p < .05) of fixations decrease with repeated reading. These main effects are qualified by interactions between repeated reading and reading experience: Shared %RECurrence (F(1, 54) = 15.49, p < .001) and shared
%DETerminism \( (F(1, 54) = 13.21, p < .001) \) decrease with repeated reading for graduate students, but increase with repeated reading for undergraduate students. Shared %LAMinarity \( (F(1, 54) = 11.44, p < .001) \) and shared ENTRopy \( (F(1, 54) = 10.32, p < .01) \) decrease likewise with repeated reading for graduate students, and increase with repeated reading for undergraduate students.

Figure 3.10. Shared %REC (upper-left), %DET (upper-right), %LAM (lower-left), and ENTR (lower-right) of fixations among graduate readers decrease with repeated reading for graduate readers, but increase for undergraduate readers. Averages are bracketed by the standard errors of the means.

For saccades, there is a marginally reliable effect of repeated reading on shared %RECurrence \( (F(1, 54) = 2.98, p = .090) \), indicating a tendency for shared %RECurrence to decline with the second reading (see Figure 3.11). Also, shared
%DETermination is reliably lower in undergraduate readers compared to graduate readers \((F(1, 54) = 10.00, p < .01)\). Finally, shared ENTRopy is also lower between undergraduate readers compared to graduate readers \((F(1, 54) = 7.90, p < .01)\), but shared ENTRopy is decreasing for graduate readers with repeated reading, while it is increasing for undergraduate readers with repeated reading, as evidenced by a reliable interaction effect between repeated reading and reading experience \((F(1, 54) = 4.43, p < .05)\). There are no other reliable effects on the recurrence quantifiers of saccades (all \(F < 2.36\)).

![Graphs showing data comparison](image)

**Figure 3.11.** Shared %REC of saccades does not reliably differ as a function of reader group or repeated reading (upper-left). %DET of saccades is higher among graduate readers compared to undergraduate readers (upper-right), %LAM of saccades does not reliably differ as a function of reader group or repeated reading (lower-left). ENTR of saccades decreased for graduate readers with repeated reading but increases for
undergraduate readers with repeated reading (lower-right). Averages are bracketed by the standard errors of the means.

Taken together, the results of cross-recurrence quantifiers of fixations suggest that there is an overall decrease of shared structure in the dynamics of fixations for graduate readers, but an increase of shared structure in the dynamics of fixations with repeated reading for undergraduate readers. For saccades, there is higher shared %DETerminism in graduate readers compared to undergraduate readers, that is, the transition from fixation to fixation follows a comparatively more ordered pattern in graduate students. Also, there is higher overall shared ENTRopy in graduate students’ saccades compared to undergraduate readers, and while shared ENTRopy decreases for graduate students with repeated reading, an increase is observed for undergraduate students. This means that graduate students’ saccades follow an overall more simple pattern which gets more complex with repeated reading, while undergraduate students’ saccades follow a more complex pattern, which gets simpler with repeated reading.

Discussion of Fixations and Saccades

Eye-movements, fixations and saccades, showed a more differentiated picture of reading performance: While over-all reading time, as measured by the average duration of fixations, also decreased with repeated reading, there was a complimentary increase in the duration of saccades. The latter finding goes together with the observed drop in the number of fixations (and saccades) with repeated reading. Monofractal analysis was again insensitive to the factors repeated reading and reading experience, whereas
multifractal analysis revealed a marginal interaction effect for fixations, with
multifractality of graduate students’ fixations showing a tendency to drop with repeated
reading, but undergraduate students’ fixations becoming more multifractal. A similar
pattern was observed for the Cross-Recurrence measures as well: Graduate students’
fixations decreased in shared structure, but undergraduate students’ fixations increase
in shared structure across all four measures. Saccades showed a more diverse pattern
of effects, highlighting differences in the heterogeneity and stability between repeated
reading and reader groups.

In regards to the fixations, we see a mostly a replication of the effects of key-
press performance observed by Wallot and Van Orden (2011a, b): The durations of
fixations decrease with repeated reading (but do not differ between graduate and
undergraduate readers). The mono-fractal structure is close to white noise, and an
interaction effect is obtained for the recurrence measures, where graduate readers’
fixations decrease in shared structure with repeated reading while undergraduates’
fixations increase in shared structure. Furthermore, we see a similar effect for
multifractality, which tends to decrease with repeated reading for graduate students’
fixations, but increases for undergraduates. With Wallot and Van Orden (2011b) it can
be concluded from the cross-recurrence analysis that undergraduate students show
stronger entrainment to the text with repeated reading. Since their fixations fall
increasingly in similar pockets of the phase-space (higher %RECurrence), the
transitions in phase-space become more similar (higher %DETerminism) and less
disrupted (higher %LAMinarity).
In contrast, graduate students’ performances individuate, share less commonality with repeated reading, but are not less stable individually, since the overall magnitudes of the individual recurrence measures are very similar. Less fluent readers increase their reading fluency in off-loading parts of their performance to general text structure, while more fluent readers follow more and more idiosyncratic patterns when fixating parts of the text, which are nonetheless constraint by the text: The closeness of the fluctuations in fixations to white noise suggests that the performance is highly constrained by the task context (see Kloos & Van Orden, 2010; Van Orden, 2010). This has also been demonstrated for changes in speed of eye-movement behavior (Wallot, Coey, Richardson, & Van Orden, 2011), but stands in contrast to the mono-fractal structure of eye-movements during visual search, which show substantial 1/f fluctuations (Aks & Sprott, 2003; Aks, Zelinsky, & Sprott, 2002; Rhodes, 2011--see also Stephen & Mirman, 2011).

This increase in multifractality of fixations with repeated reading seems to contradict the conclusions drawn from the other measures: An increase in multifractality should be an increase in the heterogeneity of variance of fixations or more frequent and more extreme changes in coupling of the processes that guide fixations (Ihlen & Vereijken, 2010; Kantelhardt et al, 2002), which could, by trend, indicate an increase in instability or at least greater heterogeneity across different scales (Mandelbrot, 1997). However, an increase in multifractality can also be gained by a decrease in variability in the laminar phases of the data. That is, the stable, low-fluctuations phases in fixations become more or stronger. However if the gaze is perturbed, maybe by a more difficult text passage, the resulting increase in fluctuations results in a more stark contrast to the
otherwise increasingly stable fixation behavior. The latter result would also be in line with the increase in shared recurrence structure observed between undergraduate readers with repeated reading.

The dynamics of saccades reveal new information about the role of reading experience and repeated reading that goes beyond what has already been observed and described elsewhere: The duration of saccades increase with repeated reading, especially for undergraduate readers. This suggests that participants make longer leaps, skipping more parts of the text between fixations. While graduate readers’ gaze already makes big leaps when reading a text for the first time, undergraduates’ increase the length of their saccades over repeated reading. At this stage, saccades change to permit a decrease of fixations, reflecting perhaps a decreased need to give the text the close reading that was necessary initially. Furthermore, saccades are very much uncorrelated, which could be—as has been explained above—due to their highly constrained nature that has been described as ballistic by some researchers (Engbert & Kliegl, 2003), although Rhodes (2011) reports that saccades during visual search exhibit strong long-range dependencies as well.

The decrease in multifractality with repeated readings also suggests that saccades are less perturbed by the text during the second reading, showing less intermittency (see Wallot et al., 2011). Together with the shorter fixations times, it seems as if participants know better how to navigate through the text, needing shorter fixation times to read the text, while the decreased fixation time suffices to specify where to look next, as evidenced by longer and also less turbulent saccades. Nevertheless, when we look at the cross-recurrence measures, we do find a difference in shared
%DETerminism for graduate and undergraduate readers, with undergraduate readers sharing trajectories in phase-space to a lesser extent than graduate readers. Given that graduate readers are indeed advanced readers, we could speculate that the difference in shared %DETerminism of saccadic movements points to the fact that graduate readers make better use of a common informational basis to navigate through the text, perhaps having identified more informative sources in the text that allow for a more effortless reading performance compared to undergraduate readers. With this interpretation in mind, the changes observed in shared ENTRopy of saccadic movements seem to suggest that the less complex (= more entropic) set of trajectories of graduate readers already reflect a high level of effortless reading, while undergraduates’ more complex eye-movements reflect a higher level of ambiguity in terms of where to aim the gaze or corrective saccades during reading of a text for the first time.

Conclusion

The replication of Wallot and Van Orden’s (2011a) self-paced reading study was only partially successful. The interaction between the reading performance of graduate student versus undergraduate students readers, where the shared dynamics of undergraduates’ performance increases with repeated reading, while the graduate readers’ shared dynamics decrease or do not change reliably, was not found in our study. However, as has been laid out this might well be due to the fact that gains in reading performance through repeated reading are lost after the 7-day between
readings period in the current study. Hence, the method of repeated readings seems to have only a transient effect on reading performance, at least with a single repeated reading session.

Nevertheless, we still observed the differences in performance between graduate and undergraduate readers, which might be due to differences in reading experience between these two groups, where graduate readers’ performance shares more dynamic structure compared to undergraduates performance in the self-paced reading task. The latter finding, i.e., that the time evolutions of graduate students’ sentence reading times showed greater commonalities compared the undergraduate, seem somewhat contradictory to the observed greater standard error in the simple Recurrence Analysis for graduate students. However, Cross-Recurrence Analysis, the analysis technique that was used to quantify commonalities between readers, basis its outcomes on the commonalities in time evolution of two series of reading times. Even if two time series are not overall more similar, but have only some certain commonalities, they will be identified by the analysis. Hence, it might be that the commonalities between graduate students’ performance are maybe intermittent, being similar over some parts of the text (maybe especially difficult or easy passages), but not uniformly throughout. Under certain circumstances, the more experiences graduate student readers might rely on a subset similar, and perhaps advantageous, reading strategies, and this is what creates the overall greater commonalities. It seems that the difference between these two groups is the strength of entrainment during text reading, where graduate readers show stronger entrainment than undergraduates, which might show that graduate readers
better exploit the text structure in the sense that they make use of it to offload control to relevant task constraints.

The eye-movement recordings corroborate this, showing similarly more shared structure between graduate readers’ dynamics in fixations compared to undergraduate readers’ for the first reading. Furthermore, fixations showed the expected drop in shared dynamic structure with repeated reading for graduate readers, and the expected increase in shared dynamics structure with repeated reading for undergraduate readers. First of all, this suggests that eye-movements prove to be a more sensitive measure than key-presses, revealing the effects of repeated reading even after a 7-day period. Second, it is interesting to note that the effects on the dynamics for key-presses and fixations are very similar. This might suggests that there is a deeper commonality that underlies the evolution of these different measures of reading and might reveal very general aspects of text reading dynamics. At least the constraints on performance by repeated reading or reading experience seem to play out similarly for these measures.

Regarding the reading performance of graduate readers and undergraduate readers, it can be concluded that the more experienced graduate readers seem to be able to exploit structure that is provided by the text initially better than undergraduates, showing more similarities in their performance. With repeated reading, however, the shared dynamics become less, and some kind of individuation of performance sets in for graduate readers, while the shared dynamics in the performance of undergraduates are increasing. It is as if undergraduate readers find improvements in reading performance by following more the structure that the text offers. It could be speculated that their performance develops with repeated reading to the point where graduate
readers started out. In the case of graduate readers, the individuation of performance might suggest that gains in reading are driven by the idiosyncratic characteristics of the reader, and rereading goes along with an increase of flexibility in performance goes along with the personal reading history of each reader, highlighting the pinnacle of reading fluency. While reading is an ‘avoidance-of-perturbation’ task for undergraduate readers, who are in search of stability, reading for graduate readers is an ‘optimization-and-fine-tuning’ task, with increases in flexible coordination at its core.
CHAPTER 4:

Study 2 - Text Difficulty and Prior Knowledge as reflected in Online Measures of Reading

Study 2 expands on findings of Wallot and Van Orden (2011a, b), as well as on the results obtained from study 1. The central questions are: First, what is the effect of text difficulty on reading performance and second, is there a simple trade-off relation between reader skill and text difficulty. In the latter case, it is expected that the differences in performance between easy and difficulty text will be the same as the difference between high and low reader skill.

This hypothesis would imply that stronger entrainment of reader’s performance to a text should be observed for reading a simple text compared to a difficult text. Concretely, measures in cross-recurrence analysis should show higher %RECurrence, %DETerminism, or %LAMinarity or all of these for reading performance of an easy text, in both, fixations and key-presses, compared to reading performance in a more difficult text.

However, increases in skill are not always just linear extension: Increases in skill might also mean that people do something not ‘just’ better, but entirely differently. For example, expert chess players seem to perceive the game and its configuration very differently from novices, memorizing it in terms of the whole configuration of pieces on the board and not piece-by-piece, which in part might be why experts are better players (Chase & Simon, 1973). Similarly, better readers might not ‘just’ be better (for example faster or possessing better memory of the text), and the interactions observed in the
patterns of fixations in study 1 hint at this fact. Hence, differences in text difficulty could also lead to surprising changes in the dynamics of performance.

While the dynamics of fixations seem to follow largely the dynamics of key-presses in self-paced reading, the dynamics of saccades provide a different kind of information about the reading process. From what was observed in study 1, it could be conjectured that more difficult text will lead to a destabilization of saccadic behavior, because text that is harder to read and comprehend will disrupt the coordination between reader and text more often.

In the following sections of this Chapter the experimental setup, procedure and materials of study 2 will be introduced. After that, the results of conventional and nonlinear analyses of key-presses and eye-movements will be presented. Finally, the findings will be summarized and discussed.

**Method**

*Design and Participants*

The experimental design features two factors: Participants read either an easy or a difficult text, and half of the participants read the text from the beginning, while the first 30% of the text were skipped for the second half of the participants (they started to read the text ‘in the middle’). Hence, the design conformed to a 2 (text difficulty: easy vs. hard/within) by 2 (prior text knowledge: reading from the beginning vs. reading from the first third on/between) factorial structure. The performance of participants was either recorded via key-presses on a keyboard or via eye-tracker.
Overall, 24 students participated in the study. Twenty-one were students of the University of Cincinnati, 3 were recruited from a non-student pool (average age 25.83 years, ranging from 20 to 35 years). All were native speakers of English and all had normal or corrected-to-normal vision. All participants read the easy and difficult texts on two separate occasions. Half \((n = 12)\) read the text by pressing a response key to reveal each new sentence in a self-paced manner, the other half read the text page-by-page, and their eye-movements during reading were recorded. Again half of each group \((n = 6)\) read the full text, and the other half read the text from the first third on. All participants were compensated in cash for their participation.

**Stimuli and Apparatus**

The texts used were the first chapters of The House Of The Scorpion by Nancy Farmer (2002) and Infinite Jest by David Foster Wallace (1996).

The first 5 chapters of The House Of The Scorpion (the easy text) describe an episode of the childhood years of Matt, a boy that grows up in the house of a servant, Celia, of a wealthy clan of drug traffickers. Matt lives a lonely life in Celia’s home, until one day kids from an estate nearby come by their house and discover him. His attempts to break out of Celia’s home, which she locks during the day to keep Matt safe inside, lead to a grave injury of Matt. The other kids take him to the estate, where he receives medical treatment, but it is also discovered that he is a clone, not a naturally borne human. Since clones are considered mere animals, Matt is locked up in the basement of the estate, while Celia and his new friends try to find a way to set him free again.
The passage from The House Of The Scorpion consists of 12,902 words, which make up 589 sentences. The average sentence length is 9.6 ($SD = 5.25$) words, and the Flesch-Kincaid index of readability assigned the text a score of 3.7, indicating that it’s difficulty is appropriate for advanced 3rd grade readers.

The first 2 chapters of Infinite Jest (the difficult text) describe the attempt of Hal, a first-class young tennis player to get enrolled into the varsity program of a North American university. The attempt fails when Hal is forced by the admission committee members to clarify some of the concerns they have regarding his application due to Hal’s emotional over-reaction in stressful social situation. The subsequent chapter describes the mental states of a person who is waiting for a delivery of Marijuana, and ruminates about the possibility that the delivery might not arrive in time or at all.

The passage from Infinite Jest consists of 12,390 words, which make up 589 sentences. The average sentence length is 20.7 ($SD = 20.19$) words, and the Flesch-Kincaid index of readability assigned the text a score of 9.3, indicating that it’s difficulty is appropriate for average 9th grade readers.

Concerns regarding the use of reading indices have been discussed in study 1. However, it is important to notice that the two texts differ in estimated readability and sentence length, both of which indicate that the texts might differ in reading difficulty for the reader.

The text presentation procedure was the same as in study 1 and so was the use of comprehension questionnaires for each story.
**Procedure**

The procedure was the same as in study 1 with one exception: After having read the stories, participants were asked to type a summary of text they had read using a keyboard and word processing software.

**Data preparation**

Data preparation was the same as in study 1, with one exception: Since Infinite Jest had considerably longer sentences that The House Of The Scorpion, no maximum response times of longer than 20,000 ms were removed from key-press responses, since this would have introduced a biased response time removal, cutting out disproportionately more reading times from the Infinite Jest text.

After preparing the data as described and calculating the key-statistics for each time series, the resulting data set was subjected to a mixed repeated measures ANOVA with the between-measures factor prior text knowledge (reading from the beginning vs. reading from the first third on) and the within-measures factor text difficulty (easy vs. difficult), separately for key-presses and eye-movements.

**Results of Key-Press Intervals**

**Averages and Standard Deviations**

Figure 4.1 presents the average duration and average standard deviation of sentences reading times as measured by key-press intervals. The average duration of sentence reading times is greater for the difficult text compared the easy text \( F(1, 10) = \)
Similarly, the average standard deviation of sentence reading times is greater for the difficult text compared to the easy text ($F(1, 10) = 73.84, p < .001$). There is no effect of the prior knowledge and no interaction between the factors (all $F < 2.63$).

Figure 4.1. Average sentence reading time is reliably lower for the easy text compared to the difficult text (left). Average standard deviation is reliably lower for the easy text compared to the difficult text (right). Averages are bracketed by the standard errors of the means.

**Mono- and Multifractal Analyses**

Mono- and Multifractal analyses were conducted. Figure 4.2 presents the results of the analyses. For monofractal coefficients, there was a reliable main effect of text difficulty ($F(1, 10) = 7.28, p < .05$), which was qualified by an interaction between the factors text difficulty and prior knowledge: While sentence reading times showed reliably more long-range dependencies for the difficult text compared to the easy text when the whole texts were presented, there was no such difference when only part of the texts were presented ($F(1, 10) = 10.38, p < .01$). There was no main effect of prior knowledge ($F < 0.24$).
Multifractality of sentence reading times was generally greater for reading times of the difficult text compared to the easy text ($F(1, 10) = 34.62, p < .001$). There was no effect of prior knowledge and no interaction between the factors (all $F < 0.38$).

Figure 4.2. Alpha exponents of sentence reading times where reliably higher for the difficult text compared to the easy text when the text was presented from the beginning. There was not difference in alpha exponents of sentence reading times between texts when the first part of the text was skipped (left). Multifractal spectrum widths were generally higher for sentence reading times of the difficult text compared to the easy text (right). Averages are bracketed by the standard errors of the means.

Surrogate analysis of monofractal estimates revealed that alpha coefficients of key-press intervals were generally reliably different from their surrogates (all $t(5) > 7.38$, all $p < .001$). Surrogate analysis of multifractal estimates revealed that key-press performance of the difficult text was reliably different from its surrogates, for whole ($t(5) = 5.65, p < .001$) and partial ($t(5) = 4.75, p < .05$) text presentation. The multifractal structure observed in key-press performance of the easy text was not reliably different from its surrogates (all $t < 0.55$).
Recurrence Quantification Analysis

The procedure to determine the delay, dimension, and radius parameter was the same as in study 1.

Figure 4.3 presents the results of recurrence analyses for %RECurrence, %DETerminism, %LAMinarity, and ENTRopy. All recurrence measures, %RECurrence ($F(1, 10) = 15.87, p < .01$), %DETerminism ($F(1, 10) = 85.81, p < .001$), %LAMinarity ($F(1, 10) = 112.92, p < .001$), and ENTRopy ($F(1, 10) = 51.15, p < .001$) were reliably higher for sentence reading times of the difficult text compared to the easy text. There were not effects of prior knowledge and no interactions between the factors (all $F < 2.32$).

Figure 4.3. %REC (upper-left), %DET (upper-right), %LAM (lower-left), and ENTR (lower-right) of sentence reading times were reliably higher for the difficult text.
compared to the easy text. Averages are bracketed by the standard errors of the means.

Cross-Recurrence Quantification Analysis

Just as in study 1, the same parameters used in simple recurrence analysis were used for cross-recurrence analysis. Since 6 data sets were available for each condition of the reading study, \( \frac{6!}{(2!)(6-2)!} = 15 \) pairing of reading time data sets were available in each condition.

Overall, the results for cross-recurrence analysis mimic those observed for simple recurrence analysis: shared %RECurrence \( (F(1, 28) = 185.71, p < .001) \), shared %DETerminism \( (F(1, 28) = 511.28, p < .001) \), shared %LAMinarity \( (F(1, 28) = 600.32, p < .001) \), and shared ENTRopy \( (F(1, 28) = 511.18, p < .001) \) are greater for the sentence reading times of the difficult text compared to the easy text.

Furthermore, a main effect of prior knowledge, equally observed for all four measures, shared %RECurrence \( (F(1, 28) = 16.17, p < .001) \), shared %DETerminism \( (F(1, 28) = 9.66, p < .01) \), shared %LAMinarity \( (F(1, 28) = 6.06, p < .05) \), and shared ENTRopy \( (F(1, 28) = 6.89, p < .05) \) indicated that commonalities in the evolution of key-presses are greater when the whole text is read compared to only part of the text.

However, these main effects are qualified by statistical interactions, again reliably observed on all four measures, shared %RECurrence \( (F(1, 28) = 15.61, p < .001) \), shared %DETerminism \( (F(1, 28) = 6.15, p < .05) \), shared %LAMinarity \( (F(1, 28) = 13.71, p < .001) \), and shared ENTRopy \( (F(1, 28) = 12.70, p < .001) \), indicating that shared structure in the temporal evolution of sentence reading times was smaller for reading
times of the difficult text when only part of it was presented compared to reading times of the difficult text when it was presented as a whole (see Figure 4.4).

Figure 4.4. Shared %REC (upper-left), %DET (upper-right), %LAM (lower-left), and ENTR (lower-right) of sentence reading times were reliably higher for the difficult text compared to the easy text, but this difference was less pronounced when the first part of the text was skipped. Averages are bracketed by the standard errors of the means.

Discussion of Key-Presses

The results obtained from key-press intervals show that reading of the difficult text took generally longer, and reading times were more dispersed as measured by standard deviation. Furthermore, long-range dependencies in response times were stronger for difficult text reading and the (cross-)recurrence measures indicated higher
(shared) structure in the temporal evolution of key-press intervals. Differences in multifractality indicated a higher degree of intermittent fluctuations in the sentence reading times of the difficult text compared to the easy text.

The effect of partial versus full text presentation was less prominent throughout the measured variables: Long-range dependencies as measured by mono-fractal analysis did not differ for key-press intervals of difficult text reading compared to easy text reading when only part of the text was presented. Furthermore, cross-recurrence analysis of key-press intervals revealed that the shared recurrent, deterministic, laminar, and complex structure was somewhat lower for partial text presentation. The difference between difficult and easy text reading was somewhat less pronounced in the partial text presentation mode.

If the interpretational framework of study 1 is adopted, then an increase in text difficulty seems to lead to an increase of entrainment, coupling the readers performance even closer to the text. As in study 1, this is evidenced by the greater shared dynamic structure in reading performance for the difficult text reading compared to easy text reading in the cross-recurrence measures. The effects seems to be particularly strong (compared to the effects that stem for the rereading and reading experience manipulations in study 1), since they are not only observed in the commonalities between reader’s performance, but also in the recurrence properties of the individual data sets, where participants’ individual reading performance also increases in stability (higher %RECurrence, %DETerminism, and %LAMinarity).

The entrainment interpretation also fits the effects of prior knowledge manipulated through the presence or absence of the initial paragraphs for each text:
There is less shared structure between participants’ performance when only part of the text was presented, and the effect is mainly due to a decrease in shared performance structure for the difficult text. It is as if the omission of prior text passages acts as a perturbation of the performance, leading to weaker entrainment of the readers’ performance to the text. Perhaps this is the case because later passages that build on the omitted prior ones lead to a decrease in text coherence, which has been identified as a factor that makes reading more difficult (McNamara, Graesser, & Louwerse, 2011).

The outcomes of mono-fractal analysis have been interpreted in two ways: One the one hand, alpha-values deviating from white noise (alpha = 0) toward 1/f noise (alpha = 1) have been suggested to indicate superior, flexible performance (e.g., Goldberger, Peng, & Lipsitz, 2002), which is due to the interconnectedness of the performance, signified by the long-range correlations in the data. Apart from the concern that reading of a difficult text should result in less superior performance compared to reading of an easy text, has come under scrutiny recently, because it is similarly at odds with many other findings of superior performance whose monofractal characteristics show a strong deviation from 1/f noise (Van Orden et al., 2011).

A different account of changes in mono-fractal structure has been offered by Kloos and Van Orden (2011), who suggest that observed increases in alpha-values from white noise toward 1/f noise are the result of a trade-off between involuntary control (such as task constraints) and voluntary control. A relative decrease in involuntary control during task performance or a relative increase in voluntary control would both lead to increases in alpha values, from white noise toward 1/f noise. Since the results of Recurrence and Cross-Recurrence Analyses seem to suggest that the reading
performance during difficult text reading is much more constrained compared to the reader performance during easy text reading, it rather seems that an increase in voluntary control is responsible for the increased alpha values for difficult text reading performance, which makes intuitive sense. The observed interaction effect between text difficulty and partial or full text presentation seems to indicate that the omission of the initial paragraphs of the easy text lead to a greater demand for voluntary control during reading and perturbing the reading performance much more compared to the difficult text. Before that conclusion can be drawn, however, one general methodological concern regarding the differences in key-press time series with different texts will have to be addressed.

As Wallot and Van Orden (2011a) reported, the lengths of sentences already determines at about 50% of the variance of key-press intervals. Since the same text was used in all conditions of study 1, the observed differences there cannot be attributed to sentence length. However, study 2 employed different texts that also differed in their sentence lengths. As can be seen in Figure 4.11, at about 65% of the variance observed in the key-press intervals was determined by sentence length. As reported in the methods sections, the two texts also differed in the average sentence length, and standard deviation of sentence length. Furthermore, a Monofractal Analysis of the ordered series of sentence lengths of the two texts yielded alpha values similar to those of the observed key-press intervals, namely 0.24 for whole text presentation and 0.30 for partial text presentation of the easy text, and 0.36 for whole text presentation and 0.28 for partial text presentation of the difficult text, respectively.
Even though the structure of the time evolution between button presses and key-press intervals is similar and thus speaks in favor of a genuine difference in terms of reading performance, high correlations between key-press intervals and sentence lengths, together with the similarities in monofractal dimension, cast doubt on the source of the effects, which could be trivially an effect of sentence length.

Figure 4.5. Correlations of sentence length with key-press intervals before (upper-left) and after (upper-right) parceling out the scaling relation between power and frequency observed in spectral analysis. Alpha-values for key-press intervals before (lower-left) and after (lower-right) parceling out the correlation between sentence length and key-press intervals. Averages are bracketed by the standard errors of the means.

To investigate this possibility, regression analysis was used to parcel out the variance explained by sentence length. The obtained residuals were again subjected to fractal and recurrence analysis. Apart from the width of the multifractal spectrum and
%RECurrence, which did not yield reliable differences between text difficulty or prior knowledge after parceling out the variance explained by sentence length, all effects observed in fractal and recurrence analysis yielded the same outcomes in the residuals that were observed in the original data, and that despite the fact that linearly parceling out of a non-time dependent structure is a considerable perturbation to fractal and recurrence analysis.

Furthermore, when spectral analysis was used to parcel out monofractal structure from the response times via Inverse-Fourier Transformation, the highly reliable correlation between sentence lengths and key-press intervals (the observed correlation coefficients where highly reliably different from zero, $t(11) = 20.60, p < .001$), was not observed any longer ($t(11) = -0.87, p = .410$).

To illustrate the opposite case, when sentence lengths were parceled out of the key-press intervals and monofractal analysis was calculated on the residuals, alpha-values were still reliably different from zero ($t(11) = 4.11, p < .01$), and a main effect of text difficulty was retained ($F(1,10) = 4.49, p = .059$ – however, no interaction effect was apparent any longer, $F < .0.04$). If the interpretation of differences in the need for voluntary control between easy and difficult text reading is kept, this result would suggest that voluntary control is more prominent throughout reading of the difficult text, irrespective of whether the texts had been presented partially or as a whole.

Similar to results on mental rotation, lexical decision and simple visual perception tasks presented by Gilden (1997), it seems that the observed nonlinear structure has priority over component effects, even when these, as in the present case, are extremely strong. However, it is interesting to note that the key-press intervals adhere to a similar
dynamic structure as was observed in the sentence lengths of the stimulus text, even when the variance of the latter was controlled for. This might suggest some kind of over-additive effect of the reading task on participants’ performance.

Results of Fixations and Saccades

Averages and Standard Deviations

Figure 4.6 presents the average duration and average standard deviations of fixations and saccades.

The average duration of fixations is reliably longer during reading of the difficult text compared to the easy text ($F(1, 10) = 6.47, p < .05$). The standard deviation of fixations is likewise greater during reading of the difficult text compared to the easy text ($F(1, 10) = 7.57, p < .05$). There is not effect of prior knowledge and no interaction between the factors (all $F < 2.26$).

For saccades, there were no reliable effects of prior knowledge or difficulty, and no interactions between the factors (all $F < 0.47$).
Figure 4.6. Average duration of fixations was reliably longer for the difficult text compared to the easy text (upper-left). Average standard deviation of fixations was reliably greater for the difficult text compared to the easy text (upper-right). Average duration of saccades (lower-left) and average standard deviation of saccades (lower-right) did not differ as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.

As can be seen in figure 4.7., there are more fixations ($F(1, 10) = 8.66, p < .05$) and saccades ($F(1, 10) = 8.65, p < .05$) observed during difficult text reading compared to easy text reading. Also, there is a trivial effect of text length, indicating that participants exhibited fewer fixations ($F(1, 10) = 5.04, p < .05$) and saccades ($F(1, 10) = 5.45, p < .05$) when the text was truncated at the beginning. There was no interaction between the factors (all $F < 0.02$).
Figure 4.7. The number of fixations (left) and saccades (right) was higher for the difficult text compared to the easy text. Averages are bracketed by the standard errors of the means.

**Mono- and Multifractal Analyses**

Figure 4.8 presents the results of Mono- and Multifractal Analyses for fixations and saccades.

Monofractal alpha coefficients are very close to white noise, but are reliably higher for fixations during difficult text reading compared to the easy text reading ($F(1, 10) = 12.16, p < .01$). There was no effect of prior knowledge and no interaction between the factors (all $F < 1.15$). There were no reliably changes in multifractality of fixations (all $F < 2.24$).

For saccades, monofractal alpha coefficients were reliably higher during difficult text reading compared to the easy text reading ($F(1, 10) = 7.26, p < .05$). Again, there was no effect of prior knowledge and no interaction between the factors (all $F < 0.29$) and there were no reliably changes in multifractality of fixations (all $F < 2.15$).
Figure 4.8. Alpha exponents were higher for the difficult text compared to the easy text for fixations (upper-left) as well as for saccades (lower-left). Multifractal spectrum widths did not reliably differ as a function of text difficulty or the presence or absence of the first 30% of the text, neither for fixations (upper-right) nor saccades (lower-right). Averages are bracketed by the standard errors of the means.

Surrogate analysis of mono-fractal estimates revealed that alpha exponents of fixations for difficult text reading were reliably different from their surrogates for whole ($t(5) = 3.67, p < .05$) and partial ($t(5) = 5.30, p < .001$) text presentation. However, alpha exponents for the easy text were not reliably different from their surrogates (all $t < 1.96$). Alpha exponents of saccades were generally not different from their surrogates (all $t < 1.77$). Surrogate analysis of multifractal estimates revealed that multifractal spectrum width of fixations all were reliably different from their surrogates (all $t(5) > 3.05, all p < .05$). The multifractal structure observed in saccades was generally not reliably different from its surrogates (all $t < 0.20$).
Recurrence Quantification Analysis

There were no effects on any of the recurrence measures for fixations (all $F < 3.17$, see Figure 4.9).

Figure 4.9. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) of fixations differed reliably as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.

Apart from a marginally reliable effect of prior knowledge ($F(1, 10) = 3.49, p = .091$), indicating a tendency for saccades to produce less turbulent fluctuations when the whole text was presented compared to only part of the text, there were no reliable effects on any of the other recurrence measures ($F < 2.79$, see Figure 4.8).
Figure 4.10. Neither %REC (upper-left), nor %DET (upper-right), or %LAM (lower-left), or ENTR (lower-right) of saccades differed reliably as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.

Cross-Recurrence Quantification Analysis

The same parameters were used for cross-recurrence quantification analysis that were estimated for recurrence analysis. Since there were 6 readers in each condition, there will be $6!/(2!*(6-2)!)=15$ paired data sets in each bin.

Cross recurrence of paired data series of fixations reveals main effects of text difficulty on all four measures, shared %RECurrence ($F(1, 28) = 20.96, p < .001$), shared %DETerminism ($F(1, 28) = 16.11, p < .001$), shared %LAM, ($F(1, 28) = 10.17, p < .01$), and shared ENTRopy ($F(1, 28) = 27.99, p < .001$), indicating that participants'}
fixations share overall more structure when reading the difficult text compared to the easy text. Furthermore, shared %RECur rence is tentatively \(F(1, 28) = 3.89, p = .059\), and shared %DETerminism \(F(1, 28) = 6.86, p < .05\) as well as shared ENTRopy \(F(1, 28) = 7.26, p < .05\) are reliably lower when only part of the text is presented compared to the whole text.

Figure 4.11. Shared %REC was higher for the difficult text compared to the easy text (upper-left). Shared %DET was higher for the difficult text compared to the easy text, and shared %DET was higher when the whole text was presented compared to when the first 30% of the text were skipped. Furthermore, when the first 30% of the text were skipped, then shared %DET was disproportionally lower for the easy text compared the the difficult text (upper-right). Shared %LAM was higher for the difficult text compared to the easy text (lower-left). Shared ENTR was higher for the difficult text compared to the easy text, and shared ENTR was higher when the whole text was presented compared to when the first 30% of the text were skipped (lower-right). Averages are bracketed by the standard errors of the means.
The two main effects of text difficulty and prior knowledge are qualified by an interaction between these two factors for %DETerminism ($F(1, 28) = 5.59, p < .05$), indicating that the drop in shared deterministic structure of fixations for the easy text is greater when only part of the text is read (see figure 4.11).

Cross-Recurrence Analysis of saccades reveals that shared %LAMinarity is generally lower when only part of the text was presented compared to when the whole text was presented ($F(1, 10) = 7.08, p < .05$). No other effects were apparent (all $F < 2.86$, see Figure 4.12).

Figure 4.12. Shared %LAM of saccades was higher when then whole text was presented compared to when the first 30% of the text were skipped. Neither shared %REC, nor shared %DET or shared ENTR differed reliably as a function of text difficulty or the presence or absence of the first 30% of the text. Averages are bracketed by the standard errors of the means.
Discussion of Fixations and Saccades

Overall, the results obtained from key-press intervals and fixations show similar patterns of effects for reading of easy and difficult texts. Reading of the difficult text took generally longer, and reading times were more dispersed as measured by standard deviation. Furthermore, long-range dependencies in response times were stronger for difficult text reading and the (cross-)recurrence measures indicated higher (shared) structure in the temporal evolution of key-press intervals and fixations across the board.

Saccades were relatively less impacted by reading difficulty: They differed only in terms of the monofractal characteristics, shifting from a slight amount of persistent fluctuations during difficult text reading to a slight amount of anti-persistent fluctuations during easy text reading.

The effect of partial versus full text presentation was again less prominent throughout the measured variables: Eye-movements also reveal only few differences between the partial and full text presentation conditions. For fixations, shared deterministic structure and shared complexity were lower for partial text presentation. The difference between difficult and easy text reading was much more pronounced in these measures compared to the full text presentation. Shared laminar structure of saccades was also generally lower during partial text presentation compared to full text presentation. This would indicate an overall weaker entrainment effect of the text on the readers’ performances, corroborating the overall picture of text reading performances from key-press performance.
Conclusion

Even though the results were not identical in terms of their effects and the absolute magnitude of the target measures for sentence-reading times as measured by key-presses and fixations of eye-movements, the degree of similarity between them seems to suggest that there might be certain fundamental commonalities in the evolution of the reading process that can be recovered with different measures.

How informative this similarity of outcomes will prove to be also hinges on whether the observed changes across these measures allow a condensation of different findings to simpler principles, and whether they allow to align the observed outcomes of different task and experimental setups. Given the results of study 1 and the commonly held assumption that reading skill and text difficulty are essentially translatable into one another (e.g., Blaxall & Willows, 1984), it would be expected that there might be a certain equality between differences in reading ability and differences in stimulus difficulty, i.e. text difficulty, as far as reading performance is concerned. This would suggest that the differences in performance between difficult and easy texts reading mimic the difference in performance between more and less skilled readers (e.g., Blaxall & Willows, 1984).

So far we have mainly focused on the role of recurrence and cross-recurrence measures to interpret the observed reading performance. However, when we base our interpretations only on theses measures, the observations from study 1 and study 2 stand in apparent contradiction to each other: While the (cross-)recurrence results obtained in study 1 (see also the results obtained for the initial reading of a text in Wallot...
and Van Orden, 2011a, b) seemed to suggest that higher reading fluency went together
with stronger entrainment of the reading performance to the text, as indicated by the
higher shared recurrence measures for graduate student readers compared to
undergraduate student readers, the results of study 2 show an increase of recurrence
measures, shared and individual (in the case of button presses), with text difficulty. It
seems as if a more difficult text would require a much closer reading, which in turn
allows readers less ‘deviation space’ in their performance. Entrainment as indicated by
increased shared temporal structure between readers’ performances would thus rather
be an indicator of reading difficulty, not facilitation. Similarly, the tendency of saccades
to exhibit persistent fluctuations and to show higher shared %LAMinarity during reading
of the difficult text compared to the easy text seems to suggest that text difficulty rather
constraints, not disrupts reading performance.

However, if the results from monofractal analysis are taken into consideration,
this apparent contradiction is resolved: As has been laid out in the discussion of key-
press intervals, changes in monofractal structure can be interpreted in terms of the
demands on involuntary versus voluntary control (Kloos & Van Orden, 2010). The
increase in long-range dependencies in reading performance with text difficulty indicate
that the ratio between involuntary and voluntary control has shifted toward an increase
of voluntary control of difficult text reading performance. Since eye-movements operate
on the level of words, entrainment to the text would mean entrainment to changes
related to the word level, which in turn leads to an increase of white noise because of
the greater involuntary constraint, but also because changes on the word level
effectively conform to a white noise process (Wallot & Van Orden, 2011a). For key-press
intervals, an increase in long-range dependencies with text difficulty was observed as well, and it was shown that this increase did not essentially depend on superficial fluctuations in sentence lengths. Hence, it seem that it is a misinterpretation to see the closer reading of difficult text as an increase in guidance by the text structure. Rather, this closer reading reflects an increase in the demand for voluntary control during reading, which is the source of the observed greater commonalities in the shared time evolution of reading times.

Finally, the few differences that have been observed for partial versus full text presentation could be interpreted as differences in prior text knowledge. They fit the overall interpretation in so far as they mimic the differences between easy and difficult text reading. However, we also observed somewhat less shared structure between readers’ performance when the initial $\approx 30\%$ of the text were not available. This could point to either a slight destabilization of reading performance every time when prior knowledge would be necessary for text comprehension but is not available. Alternatively, lack of prior knowledge could lead to an individuation of performance in the sense that readers are forced to make up prior knowledge though interpretations of the text to maintain comprehension. Since a common base for these interpretations is smaller when the initial part of the text was missing, this would increase the degree of idiosyncrasies in text comprehension, and this is what is observed as a drop in the shared dynamics during reading. However, this does not seem to have a great detrimental effect on reading performance.
CHAPTER 6:  
General Discussion

The presented studies were an attempt to shed light on the role of reading skill, text difficulty, and prior knowledge on text reading performance as measured by key-presses and eye-movements. Complexity metrics were applied to gauge the dynamics of the reading process. The holistic quantification of task performance made possible by these measures seem to be an appropriate, perhaps even necessary, next step for the progress of reading research, since the observed stimulus-response interactions are not likely reducible to a few component effects. That is, the psychologically salient properties of words, sentences, or texts are fundamentally context dependent, playing different roles in different contexts and do not always lead to the same outcomes. In fact, even drastic simplifications of the reading process, such as single-word reading studies, have eluded the goal of reduction to a simple performance structure.

The promise of new metrics to evaluate performance complexity is to cut this Gordian knot and take a new perspective on reading. The presented studies have shown that diverse manipulations of text reading result in reliable changes of the holistic structure of reading dynamics. However, two challenges stand between the current findings and an ultimately productive usage of complexity metrics in reading research. One is to broaden the empirical basis for complex text reading characteristics, employing a wider variety of texts, measures, participants and languages. Any conclusions that can be drawn from the current research are necessarily limited to the specifics of the presently small data basis and cannot circumvent the obvious fact that
this makes the present two studies a mere candle in the dark. This brings about the second challenge, which is to present an appropriate interpretation of the outcomes of the presented studies with regard to the concrete results themselves, but in a manner that does justice to the phenomenon at hand and the new perspective, the new language, that is introduced by the framework of complexity science.

The latter challenge is addressed in the following: Firstly, it is discussed what has been learned regarding reading skill, text difficulty, and prior knowledge. Secondly, this discussion is expanded in two directions, addressing the differences and similarities of the outcomes between the two types of measures, key-press intervals and eye-movements, as well as discussing and criticizing the notion of ‘entrainment’, which as been put forward as a leading construct to interpret complex text reading performance (Wallot & Van Orden, 2011a). Furthermore, the observation and meaning of individual differences in reading and its relation to stimulus materials and measurement procedures are discussed. Finally, the general role of nonlinearities in complex task performances is outlined; and an outlook for the development of reading research concludes this dissertation.

Reading Skill, Text Difficulty, and Prior Knowledge

The ultimate outcome of differences in reading skills are differences of text comprehension and understanding (Clifton & Duffner, 2001; Crowder & Wagner, 1992; Rayner & Pollatsek, 1994). Comprehension is the gold standard by which the outcome of a reading performance should be evaluated. As true as this statement sounds, as
complicated in a reality it is embedded: Even if we ignore the problem of how to measure comprehension appropriately (i.e., what does it take to address comprehension? Is comprehension sufficiently captured by word insertion or recall tasks?--see Rayner and Pollatsek, 1994), readers can make up for lack in skill with effort or prior knowledge (McKeown, Beck, Sinatra, & Loxterman, 1992; O'Reilly, & McNamara, 2007). Also, comprehension seems to be amendable to a variety of context effects, which are not in a narrow sense part of the reader or the text (e.g., Bohn-Gettler & Rapp, 2011), but the task situation as a whole. The interrelatedness between these factors, as well as the flexibility of readers' performances makes the occurrence of simple trade-off relations between them an exception (Wallot & Van Orden, 2011c). The holistic analysis of performance, focusing on the organization of performance instead of the specific performance outcomes might reveal such simple trade-off relations by taking into account the nonlinearities that are involved in producing them. However, the results of the two presented studies do not lend themselves to such an interpretation easily.

The studies by Wallot and Van Orden (2011a, b) that investigated text re-reading performance of readers of different reading fluency suggested that (very) skilled reading is akin to some sort of entrainment process, whereby the reader's performance, as measured by key-presses, entrains to the text structure during reading, perhaps offloading performance control to the structure of the text. The basis for this interpretation was mainly the observed differences in recurrence, deterministic and laminar structure between readers of different skills and text knowledge: While über-fluent graduate student readers showed more recurrences, deterministic and laminar...
structure in sentence reading times, less fluent undergraduate readers showed less of these structural indices. Furthermore, the recurrence patterns of graduate readers remained fairly stable across the first and second reading of the same text or show a trend to decrease with repeated reading, while undergraduate readers’ reading times became increasingly recurrent, deterministic and laminar, showing an increase in entrainment. Differences in the fractal properties between graduate and undergraduate readers, and between readers who read the text for the first or second time were not apparent.

In study 1, we observed similar differences between graduate and undergraduate student readers (although without the reliable increase in undergraduate readers’ performance) in key-press intervals, and we observed the same differences in fixations between graduate and undergraduate readers for the first reading, as well as the increase in structure in undergraduate readers’ performance. Even if the drop in shared structure of graduate readers’ fixations with repeated reading is ignored, the prediction that a more favorable relation between capabilities on the side of the participant and difficulties on the side of the stimulus text should be observed as entrainment, indicated by higher values on the structural measures in recurrence analysis, did not hold for study 2. Here, key-press intervals and fixations showed increased structure, as evidenced by %RECurrence, %DETerminism, and %LAMinarity for the difficult text compared to the easy text.

While it remains speculation how participants’ reading performance for the difficult text would have changed with repeated reading, the observed monofractal structure captured a difference between easy and difficult text reading: The reading of the difficult
text resulted in higher alpha values (i.e., stronger long-range dependencies) compared to the relatively easy texts, for key-press intervals and fixations. This is in line with the account of changes in monofractal structure in human performance offered by Kloos and Van Orden (2010; see also Van Orden 2010), who describe the observation of fractional gaussian noise as due to a trade-off between involuntary and voluntary control, where decreases in involuntary control and increases in voluntary control lead to higher alpha-values (that is, stronger long-range dependencies), while increases in involuntary control and decreases in voluntary control lead to lower-alpha values (that is, weaker long-range dependencies). The higher alpha values observed in the reading performance of the difficult text thus indicate an increased demand for voluntary control on the side of the reader, which seems a plausible interpretation.

As has been discussed, it seem very plausible that the increase in shared dynamics with increased text difficulty captures similarities in the performance that are not due to the text structure, but due to the overarching organization of the reading performance given that text structure (i.e. that of the difficult text). With regard to the current body of knowledge on the dynamics of text reading, this is a satisfying interpretation, and this interpretation is also corroborated by the observed increase in the %LAMinarity and ENTRopy measures with text difficulty that suggest a certain homogenization of performance, while a difficult text should rather result in a heterogenous performance, as far as the text structure is regarded. The reader is working ‘against’ the text, so to speak.

However, this points also to the limitation of structural recurrence measures, which have been the hallmark-measures to assess entrainment in Wallot and Van Orden
(2011a, b). We will return to the question of entrainment in the next section of the general discussion. Before that, we have to ask how the findings regarding prior text knowledge do fit into this picture?

By large, they seem to simply mimic the differences between fully presented easy and difficult texts with the exception of the results obtained from cross-recurrence analysis of fixations, where fixations of readers who read only part of the story shared less structure compared to the fixations of readers who read the whole text. Furthermore, the difference between the amount of commonalities in the difficult reading condition and the easy reading condition was amplified as well, compared to the full text condition. This suggests two things: First of all, there is evidence that lack of prior knowledge changes reading performance in the sense that it acts as a perturbation to the reading process or as an amplification of individual differences. Since the evolution of the text is what constraints performance, if part of that text is missing, then it should follow that readers’ performance is somewhat less constrained. That the overall analyses outcomes of the partial text presentation are still very similar to the whole text presentation condition might simply be a consequence of reading being much more of a constructive process. Participants seek meaning in what they read, and even if a text is more ambiguous when parts of it are missing, this does not prevent the reader from eventually catching up on the content—or some content. After all, when reading researchers select text stimuli, these are most often only text passages, lacking other parts of the text they are taken from (e.g., LeVasseur, Macaruso, Palumbo, & Shankweiler, 2006; LeVasseur, Macaruso, & Shankweiler, 2008; Just & Carpenter, 1980).
The interaction effect observed between text difficulty and prior knowledge, even though small in magnitude, fits in this picture: Since the easier text follows a much more linear mode of story telling, reading performance is disproportionally impacted by the absence of the first parts of the text, since the unfolding story builds upon it in a very straight forward manner. For the difficult text, this is not the case: The mode of story telling is more complex, also more ambiguous, with many sidetracks. The absence of the first passages is in so far much less of a loss, since the text is never straight forward, and requires a certain constant interpretational effort, which does not get resolved by the presence of a particular piece of information.

**Entrainment as the Core-Concept of Text Reading**

Entrainment as core concept for the dynamics of text reading, even though it seems to capture some of the salient aspects of text reading well, is problematic with regard to at least two aspects: First, as the current research shows that it is difficult to assess entrainment with one measure or a small set of measures, such as the commonalities in the evolution of measurement series of the reading process. They rather seem to be ambiguous in their relation to variations of change in difficulty of a reading task, depending on other measures, such as the observed fractal structure, to interpret them. Second, the notion of entrainment still assigns a certain priority to the text, following the lead of classical stimulus-response theories. Entrainment always assigns a hierarchy between a leader and a follower, the performance of the reader following the structure of the text. Reading, however, is not such an activity. Rather,
reading is a coordination between reader and text. The notion of synergy is more fitting, since no text can be evaluated without the characteristics of the reader, and the reader’s characteristics factor into the characteristics of the text.

However, the real test of a synergetic relationship would be the perturbation of the text-reader system at points of different stability (Haken & Schiepek, 2006; Kelso, 1995), which would require more advanced experimental procedures than those used in the present study. Perturbations of the reading process could be produced by inserting sentences into a text that do not belong there, or locally reduce the coherence of a text. Measures with high sampling frequency, such as eye-movements, could then be used to evaluate the effect of such perturbations in different textual contexts.

On the positive side, the interpretation of reading as a synergistic activity might have direct repercussions for pressing questions in applied reading research, as it provides a framework to test and evaluate the stability and flexibility of reading performance of a participant; for example, the magnitude of critical slowing down after the perturbation of reading performance. Furthermore, factors that promote stability in reading performance might be a straightforward target measure to evaluate intervention techniques to improve the reading performance of any population of interest. For now, this has to remain speculation. However, the belief in the presence of a fairly encompassing synergy that is the basis of observed reading performance is further strengthened when the commonalities between the different employed measures of reading performance are considered.
Eye-Movement and Key-Press Measures of Reading Performance

Even though there are differences in the dynamics between reading performance when measured by key-press intervals compared to eye-movements, especially fixations, the commonalities are more striking: While fixations do not mimic the effects observed in key-press intervals in study 1, they do replicate the findings presented in Wallot and Van Orden (2011a, b). For study 2, the effects and their directions for Monofractal Analyses, as well as for Cross-Recurrence Analysis, are very similar; and this despite the fact that key-presses are not only a much more coarse-grained measure of reading performance, but are also highly influenced by the mere physical length of a sentence.

This makes it seem as if the synergy or synergies that underlie text reading stretch out from the scale of milliseconds to the scale of several seconds, from eye-movements to key presses. On the other hand, while in study 1 eye-movements seem to be a more fine-grained measure, capturing effects of repeated reading that the key-press intervals could not, the results of study 2 give the impression that key-press intervals capture the differences between easy and difficult text reading more reliably, as evidenced by the observed effects in simple Recurrence Analysis. In some sense, it seems that there is a certain complimentarily at work in the eye-movement measures, that allows a more fine-grained resolution of the reading process on the one had, but on the other hand also allows a greater expression of individual differences.

The latter point might also prove to be a good interpretational framework to make sense of the observed interaction effect of partial versus whole text presentation in
cross-recurrences of fixations in study 2: Maybe the performance in the partial text presentation is not less stable or has been perturbed by the omission of the initial text passages, but forces individual differences on the side of the reader to ‘fill the gaps’. Each individual performance in the partial text presentation condition might be as stable as one from a reader in the full text presentation condition, but the smaller common basis leads to more different performances. That this effect is pronounced for performance during easy text reading might—again—be due to the increased need on the side of the reader to form more detailed interpretations of what might have happened earlier in the story to allow for a coherent understanding of the story, which in turn would be drawn from reader specific properties and perhaps be more amendable to contextual effects (e.g., Bohn-Gettler & Rapp, 2011).

**Conclusion and Outlook**

The presented studies make a strong case for the use nonlinear metrics in reading research. Recurrence and Fractal Analyses distinguished reliably between readers with different levels of reading experience, and reading of easier and more difficult texts. These outcomes promise to be a basis for the investigation of reading as a process during complex reading tasks. The similarity of outcomes observed between key-press intervals and eye-movement recordings suggest that a certain degree of commonality in outcomes across measures can be expected.

Furthermore, the observed nonlinear structure does not only offer information about the reading process, but also trumps very strong linear correlations in sentence
reading times as measured by key-press intervals. That is, the--in the case of sentence
length trivial--linear dependencies observed in reading task performance is couched into
nonlinearities, which in turn hold information about the relevant differences between
task manipulations.

What fundamentally distinguishes more fluent from less fluent readers or easy
reading tasks from difficult ones remains yet to be discovered. The capacity to entrain to
a text seems to be a good starting point to make sense of the observed differences in
performance. Further, the presented results suggest that the difference will likely center
around the notions of stability and flexibility. Some of the next steps that should be
taken to answer this question have already been discussed and include the sampling of
a more diverse population of readers and texts to broaden the general empirical basis of
investigations into the dynamics of text reading. Furthermore, to test the synergy
interpretation that has been put forward in this dissertation, more sophisticated
perturbation studies in which readers’ performance is disrupted and the effects of these
disruptions are carefully studied are to be conducted. The reward might be an even
richer picture of reading performance. Obviously, eye-movement recording would be the
method of choice.

Whether or not any of these conceptions will stand the test of time, the outlined
approach clearly opens up new avenues to the investigation of reading that do not
merely have the genuine relation between text and reader as a shimmer on the horizon
that remains outside of the scope of conductible research, but as the center piece of the
science of reading.
References


Fisher, D. F., & Shebilske, W. L. (1985). There is more that meets the eye than the eyemind assumption. In R. Groner, G. W. McConkie, & C. Menz (Eds.), *Eye movements and human information processing* (pp. 149-158). Amsterdam: North Holland.


