Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 6, 1745–1765

0278-7393/14/$12.00 http://dx.doi.org/10.1037/xlm0000030

The Role of Reading Time Complexity and Reading Speed in Text Comprehension

Sebastian Wallot
Aarhus University

Beth A. O'Brien
National Institute of Education

Anna Haussmann and Heidi Kloos
University of Cincinnati

Marlene S. Lyby
Aarhus University

Reading speed is commonly used as an index of reading fluency. However, reading speed is not a consistent predictor of text comprehension, when speed and comprehension are measured on the same text within the same reader. This might be due to the somewhat ambiguous nature of reading speed, which is sometimes regarded as a feature of the reading process, and sometimes as a product of that process. We argue that both reading speed and comprehension should be seen as the result of the reading process, and that the process of fluent text reading can instead be described by complexity metrics that quantify aspects of the stability of the reading process. In this article, we introduce complexity metrics in the context of reading and apply them to data from a self-paced reading study. In this study, children and adults read a text silently or aloud and answered comprehension questions after reading. Our results show that recurrence metrics that quantify the degree of temporal structure in reading times yield better prediction of text comprehension compared to reading speed. However, the results for fractal metrics are less clear. Furthermore, prediction of text comprehension is generally strongest and most consistent across silent and oral reading when comprehension scores are normalized by reading speed. Analyses of word length and word frequency indicate that the observed complexity in reading times is not a simple function of the lexical properties of the text, suggesting that text reading might work differently compared to reading of isolated word or sentences.

Keywords: text reading, comprehension, reading speed, recurrence analysis, fractal analysis

Skilled reading is associated with good comprehension. Comprehension of a text arises during the reading process, and skill shapes this process so that better comprehension can be achieved—or so that the same level of comprehension can be achieved with less effort and in shorter time. However, relating the process of reading to its outcome, that is, relating the degree of skill exerted during reading to the understanding of a text, has proved to be anything but trivial. The main aspect of the reading process that has been investigated in the literature is its "fluency." Fluency is usually operationalized as reading speed, given that comprehension is sufficient. This is spelled out in theories of automatic and fluent reading processes, where automaticity in word identification (e.g., LaBerge & Samuels, 1974) trades off with the effort to comprehend text (e.g., verbal efficiency theory; Perfetti, 1985). Thus, fluent word reading should contribute to better comprehension, and vice versa. Accordingly, studies report high correlations between measures of the speed aspect of reading fluency and comprehension (Hosp & Fuchs, 2005; Hintze, Callahan, Matthews, Williams, & Tobin, 2002; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003), and it is believed that this relation is a bidirectional one (Dowhower, 1987; Fuchs, Fuchs, Hosp, & Jenkins, 2001).

However, a conceptual problem arises in this context. On the one hand, reading speed is used as a descriptor of the reading process. It is considered an index of skill such that faster reading indicates fluency and correlates positively with comprehension. Measures of timed reading that use words-per-minute reading rates to score reading ability, such as the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2002) and the Gray Oral Reading Test (Wiederholt & Bryant, 2001), rely on reading speed in this way to arrive at a reading score. On the other hand, reading speed can be considered as an outcome of the reading process itself (Kintsch, 1998; Kuhn & Stahl, 2003). Furthermore, reading speed

This article was published Online First July 7, 2014.

Sebastian Wallot, Aarhus University; Beth A. O'Brien, Centre for Research in Pedagogy and Practice, National Institute of Education; Anna Haussmann and Heidi Kloos, Department of Psychology, University of Cincinnati; Marlene S. Lyby, Interacting Minds Centre and Unit for Psychooncology and Health Psychology, Aarhus University.

Preparation of this article was supported by grants from the National Science Foundation to Guy Van Orden (Grant NSF BCS 0642716 and NSF BCS 0843133), and to Guy Van Orden and Heidi Kloos (NSF DHB 0728743). Sebastian Wallot acknowledges funding by the Marie-Curie Initial Training Network, TESIS: Towards an Embodied Science of Inter-Subjectivity (FP7-PEOPLE-2010-ITN 264828).

Correspondence concerning this article should be addressed to Sebastian Wallot, Interacting Minds Centre, Aarhus University, Jens Chr. Skous Vej 4, Building 1.483, 3rd floor, 8000 Aarhus C, Denmark. E-mail: sewa@hum.au.dk

as a process measure has a natural correlation with reading skill, but can also be used to compensate for the lack thereof—for example, when readers trade in reading time to increase comprehension (Carver, 1992). In this way, adjusting one's reading rate is a strategic resource, and the relationship between reading speed and comprehension has remained elusive when measured on the same text—sometimes showing positive correlations, but more often yielding null results (compare Schroeder, 2011, and Berninger et al., 2010, with LeVasseur, Macaruso, & Shankweiler, 2008; LeVasseur, Macaruso, Palumbo, & Shankweiler, 2006; McNerney, Goodwin, & Radvansky, 2011; Wallot, 2011b). The ambiguous nature of reading speed—being both process and product, skill and compensatory resource—requires a careful examination of how we conceptualize and quantify the reading process.

The goal of this article is to resolve these conceptual issues and diverse findings by presenting a different conceptual and measurement perspective on the reading process. As an alternative to operationalizing reading fluency as speed, one can consider reading fluency in terms of the structuredness, stability, or instability of the reading process (Wallot & Van Orden, 2011). As readers progress through a text, their understanding develops (i.e., Donald, 2007; van Dijk & Kintsch, 1983), and it has been suggested that reading can be seen as a dynamic system, with fluctuations and phase transitions and attractors that are shaped by the quest for meaning (Paulson, 2005). A reading process that brings about good comprehension can be understood as being subject to constraints, as if being pulled into a specific semantic/conceptual space that delimits options for interpretation of what is being read (e.g., Frazier et al., 2005). As long as readers stay in this semantic subspace and maintain good comprehension, the reading process proceeds smoothly. As comprehension is jeopardized, the reading process is perturbed, for example, as evident in sudden increases in reading times within garden path sentences (Bailey & Ferreira, 2003), or the initiation of regressive eye movements during reading of inconsistent sentences (Rayner, Chace, Slattery, & Ashby, 2006).

The problem with these examples is that they confound the perturbation of the successful reading process with its speed (or lack thereof). What is needed is measures of the reading process that quantify the degree of perturbation independently from speed. From a measurement perspective, information about the temporal order within the reading process could be a solution to the problem. Time series analysis techniques drawn from complex systems theory, which are now applied to data from physiology and psychology, might provide a potential solution here (Wallot & Van Orden, 2011). Moreover, applying ideas from the framework of complex systems to reading provides new conceptual ideas for how to understand the reading process. Complexity metrics characterize cognitive performance in terms of structure, instability, and interdependence, and it has been argued that these properties are indicative of cognitive organization during tasks (Holden, Van Orden, & Turvey, 2009; Van Orden, Holden, & Turvey, 2003).

Three measures that capture these aspects of reading times are monofractal scaling, multifractal scaling, and percent determinism (%Determinism) of recurrence plots, which are described in the following sections. By employing these measures, we ultimately seek a means of measuring reading processes related to comprehension that does not rely solely on speed, which obfuscate process versus product, and skill versus compensatory resource. Rather, measures are called for that capture the process capabilities of comprehension and how the reading process adapts to changes in comprehension.

In a previous article, we examined reading speed together with two of these three measures of reading times: monofractal scaling and recurrence (O'Brien, Wallot, Haussmann, & Kloos, 2013). The aim was to see whether these measures distinguished between readers of different age (Grades 2–6 and adults), where age served as a proxy for reading skill. We found that reading speed and recurrence measures distinguished groups of readers of different ages, but the monofractal scaling measure did not show differences between the groups.

In this article we present a reanalysis of that data in order to systematically expand the previous, more exploratory work: Instead of examining gross differences in reading skill between groups, we investigate how individual differences in concrete aspects of reading skill, namely, speed and comprehension, can be predicted by complexity measures of the reading process. This is of importance, as especially the prediction of text comprehension from features of the reading process has proved to be a persisting challenge (Berninger et al., 2010; LeVasseur et al., 2006, 2008; McNerney et al., 2011; Wallot, 2011b). Also, since monofractal scaling did not yield reliable differences between reader groups (O'Brien et al., 2013), we include an analysis of multifractal scaling in reading times, which is an extension of monofractal scaling, to investigate whether the time dependency of scaling might have masked effects of the overall strength of scaling (Mandelbrot, 1997). Finally, we also conduct an analysis of the role of lexical variables in this reading task. This is important in order to clarify the origin of complexity properties in reading times. One the one hand, lexical variables are thought to be a main driving force behind the reading process, and corpus analyses showed that the distributions of lexical variables in connected texts also exhibit complex fractal (Ebeling & Poeschel, 1994; Montemurro & Pury, 2002) and recurrence (Orsucci et al., 2006) properties that we use in our analysis of reading times. On the other hand, recent work on text reading found that lexical features do not play a substantial role in connected text reading (Wallot, Hollis, & van Rooij, 2013).

Before we present the details of the study and the results of the analyses, the aim of the next sections is to give a brief description of what the applied complexity metrics mean statistically, what their interpretation is in the context of text reading, and what their putative relations to comprehension are.

Monofractal Scaling

Monofractal scaling is a property of a time series that indicates the strength of power–law scaling relations, also called long-range correlations, across a series of measurement values. Trial-by-trial measurements, such as consecutive response times obtained from an experiment, differ in terms of the degree of temporal carryover effects from one trial to the next. If response times are strictly independent of each other, then there is no carryover effect, and the response time fluctuations are random, conforming to a white noise pattern. If response times possess short-term correlations, then local carryover effects are observed, whereby the response on trial t_2 is affected by the outcome of trial t_1 and somewhat affected by the outcome of trial t_0. Here correlations between trials decay quickly, affecting only—more or less—adjacent trials. However, in recent years monofractal scaling relations have been reported in many psychological experiments, where carryover effects span the whole sequence of trials in an experiment (see Kello et al., 2010; Van Orden, Kloos, & Wallot, 2011). This means that trial responses are not just a function of the processes on fast timescales that connect one trial's onset with the execution of the response on that particular trial, but that processes across many timescales (and trials) interact to produce a particular trial response.

The presence of scaling relations is usually observed as a pattern of fluctuations over time, whereby the fast timescales (i.e., fluctuations across only a few trials) are small in magnitude, but are nested in fluctuations on slower and slower timescales that increase in magnitude. The magnitude of fluctuations increases proportionally with the size of the timescale, and the faster the fluctuations grow with increasing timescale, the stronger the effect of long-range dependencies in the time series (see Figure 1).

An early hypothesis for the role of monofractal scaling in cognitive performance was that optimal performance in a task yields a so-called pink noise pattern. It was thought that pink noise would indicate skill and health of the cognitive systems (e.g., Goldberger et al., 2002; Wijnants, Bosman, Hasselman, Cox, & Van Orden, 2009). However, there have been several inconsistent findings with regard to the role of pink noise as an indicator of optimal performance (˜ skill), and it has been proposed that strengths of scaling relations are rather indicative of the degree to which external control sources drive cognitive performance (Kloos & Van Orden, 2010).

For example, the clearest cases of monofractal scaling have been observed in response time tasks under stable task conditions, such as simple reaction times (Holden, Choi, Amazeen, & Van Orden, 2011), tapping (Chen, Ding, & Kelso, 1997), or time estimation (Gilden, 2001). The strength of scaling relations can also distinguish between different experimental conditions or tasks. For example, Kuznetsov and Wallot (2011) asked participants in a time estimation task to press a button every time a second had passed. Participants were provided either with accuracy feedback about their performance or with no feedback at all. In the no-feedback condition, they produced time estimates with strong scaling, that is, highly interdependent response times. In the feedback condition, scaling strength was significantly reduced; that is, their time estimates were constrained to more local, within-trial dependent fluctuations. That is, when participants processed information from the environment to perform the time estimation task, long-range dependencies in their time series' data decreased and therefore monofractal scaling also decreased. This interpretation is also in line with earlier findings in a study on silent versus oral reading, where oral reading produced reduced monofractal scaling, as articulatory processes during reading enhanced the effect of local word properties on individual reading times (O'Brien, et al., 2013).

In the case of reading, it is usually assumed that lexical or other text properties are the driving informational force of reading times (e.g., Coltheart, Rastle, Conrad, Langdon, & Ziegler, 2001; Engbert, Nuthmann, Richter, & Kliegl, 2005; Grainger & Jacobs, 1996; Reichle, Rayner, & Pollatsek, 2003), affecting the length of

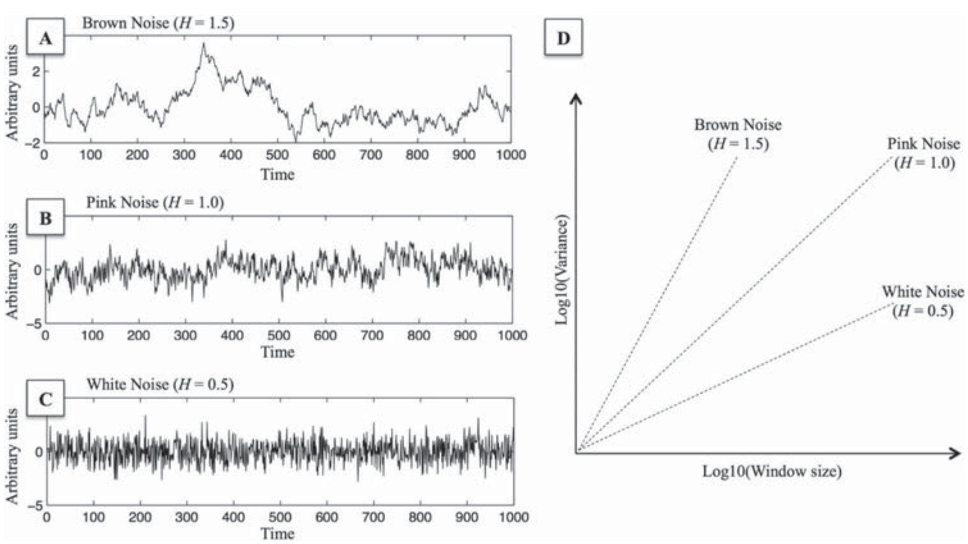

Figure 1. Illustration of monofractal time series and their monofractal scaling properties in detrended fluctuation analysis (DFA). (A) Time series of idealized brown noise. (B) Time series of idealized pink noise. (C) Time series of idealized white noise. (D) The DFA log-log plot illustrates the change of variance with changing window size for the three noises. As can be seen, variance increases faster with growing sample size as the time series deviates from white noise toward brown noise. When the logarithm of the variance is plotted against the logarithm of the sample size over many different sample sizes, the slope of the line that captures the relation between variance and sample size gives the Hurst exponent H, which estimates the strength of monofractal fluctuations in a time series. A Hurst exponent of $H = 0.5$ indicates growth of variance with increasing sample size in line with the central limit theorem, which suggests independence of data points in the time series. Deviations of H toward greater numbers indicate greater degrees of interdependence of data points in the time series.

fixations and response times during reading (Reingold, Reichle, Glaholt, & Sheridan, 2012).

Just as in the example of time estimation from Kuznetsov and Wallot (2011), when repeated keypresses in time estimation are guided by external feedback, keypresses that are contingent on text properties should also lead to a decrease of long-range correlations in the series of time estimates. Hence, we hypothesize that skilled reading (i.e., high reading speed and high comprehension) should be indexed by reading times that are systematically driven by text features, as indicated by weak traces of long-memory across reading times, that is, reading times that are closer to white noise.

Multifractal Scaling

Multifractal scaling is an expansion of the concept of monofractal scaling. If a time series, for example, of reading times, exhibits significant multifractal structure, reading times are very heterogeneously distributed in the time series, with periods of low-magnitude fluctuation interspersed with periods of high-magnitude fluctuation. The occurrences of these different fluctuation patterns over time are not random but correlated with each other. This distinguishes multifractal fluctuations from simple outliers, that is, the random occurrence of isolated data points that "stick out" of the overall pattern of measurements. Hence, multifractal fluctuations signify systematic fluctuations that are indicative of interactions among timescales (Ihlen & Vereijken, 2010; Kelty-Stephen, Palatinus, Saltzman, & Dixon, 2013).

Statistically, these instabilities can be seen as a change in local monofractal patterns (see Figure 2). As laid out, monofractal structure in reading times can be interpreted as the degree of constraints that are brought by the text structure onto the intervals between button presses, turning them into measures of word reading times. Multifractals quantify the degree to which these constraints change over time and lead to on-line reorganizations of the reading process during text reading—the breaking and formation of cognitive constraints (Stephen, Anastas, & Dixon, 2012).

Even though multifractal fluctuations are observed throughout a whole variety of cognitive measures (Ihlen & Vereijken, 2010; Stephen et al., 2012), no systematic experimental manipulations have been devised that change mulitfractal structure in cognitive tasks. Still, multifractal structure has a special conceptual appeal for reading, as it is indicative of sudden on-line changes that occur in a time series of measurement values. We argue that such changes capture an adaptive aspect of reading fluency that indexes drastic changes in comprehension on the side of the participant.

In contrast to this, current theories of the reading process that try to capture the emergence of meaning above the level of word features assume the process of text understanding to be a piecewise buildup of information, adding bits of information as the reader advances through the text (e.g., Donald, 2007; Zwaan, Magliano, & Graesser, 1995). Here comprehension is conceptualized as a gradual process akin to building a mosaic piece by piece. In some circumstances this seems plausible. However, the development of comprehension during reading may involve more than just gradual shifts. In a study by Zwaan et al. (1995), situation model variables explain only tiny amounts of variance in text reading times, and in McNerney et al.'s (2011) study, in which participants read a whole book, the situation model approach comes to its limits: From a larger set of predictor variables that should code for psychologically relevant content, some show the expected effects, some yield null results, and some show effects exactly opposite to the predictions. Perhaps this could be a function of comprehension that does not always evolve in a gradual way. An example of how little pieces of information can dramatically change understanding is provided in Asch's (1946) classical study. Participants were presented with different lists of adjectives that described a putative person. Afterward, participants were asked to characterize this person. While the presence or absence of some of the adjectives on those lists (such as *intelligent* and *determined*) led to only minor changes in the characterizations made by the participants, other adjectives (such as *cold* and *warm*)

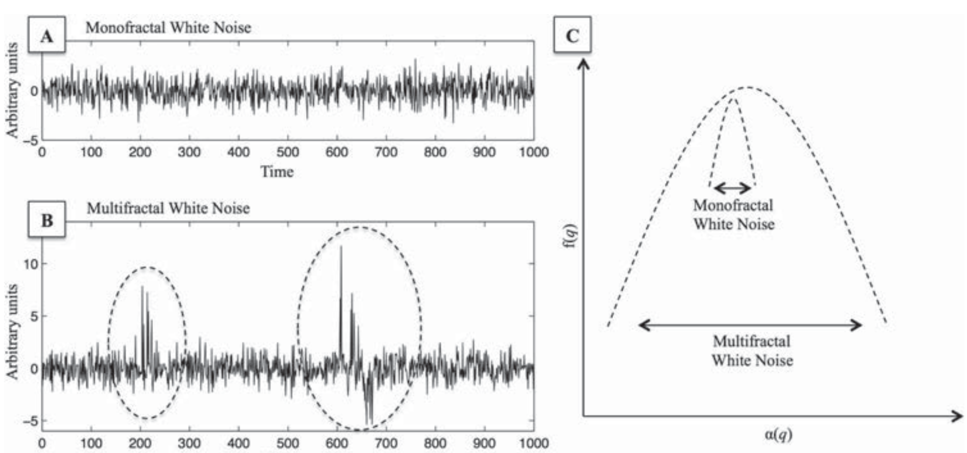

Figure 2. Monofractal and multifractal fluctuations. (A) Monofractal white noise time series. (B) Multifractal white noise time series. As can be seen, the multifractal time series displays intermittent fluctuations, suggesting interactions among timescales. (C) The multifractal spectrum width is the difference between the smallest and largest $\alpha(q)$ values. The bigger the width of the multifractal spectrum, the more heterogeneous the fractal scaling properties of a time series.

completely changed the impression of the putative person and also changed the meaning of the other adjectives on the list.

Multifractal fluctuations might pick out exactly these kinds of pivotal conceptual changes in the fluctuation patterns of reading times. Along this line of reasoning, we hypothesize that the presence of multifractal structure is an index of adaptive development of comprehension during reading, quantifying the degree to which the reader picks up important pieces of information in a text and makes appropriate connections within the larger context of the text's content. However, this is a speculative proposal, given the few studies that have examined this property in cognitive tasks (Stephen et al., 2012).

%Determinism of Recurrence Plots

%Determinism is a measure of the stability and structuredness of fluctuations in a time series, obtained from recurrence quantification analysis (RQA). RQA is a method to characterize the behavior of complex time series that are the result of many interacting variables (Webber & Zbilut, 1994). Text reading is a complex cognitive activity that involves perceptual, memory, reasoning, and output processes (Reichle & Reingold, 2013; Schroeder, 2011; Van Orden, Holden, Podgornik, & Aitchison, 1999). Although all these processes contribute to the product of reading, we are limited to a single measurement of reading (i.e., the one-dimensional time series of word reading times in the present study), and therefore our measurement conflates the contribution of these different processes. The method of phase space reconstruction through time-delayed embedding (Takens, 1981) can be used to work around this problem of not having independent measures of all contributing processes. By utilizing the method of phase space reconstruction, RQA allows for the reconstruction of the multidimensional dynamics of the cognitive system from a single, one-dimensional time series. Phase space construction is possible because the single time series of measurements contains information from all participating processes. Figure 3 gives an example of phase space reconstruction by using time-delayed copies of the original time series as surrogates for the dimensions of the reading dynamics that have not been directly measured.

The manner in which a phase space is filled by the time-delayed signal is called an attractor. As can be seen in Figure 3, the trajectories in the example phase space show a clear, repetitive structure, crossing the same neighborhoods in the phase space again and again. This behavior of the time series converts to diagonal lines on the recurrence plot (see Figure 3C) and signifies a high degree of stability of the temporal behavior. The measure of %Determinism is a statistic of stability and structuredness of the time series that is created by quantifying the extent to which a time series produces such diagonal line structure—repetitive sequences—over time.

RQA has been applied in many psychological tasks, mainly to assess aspects of interpersonal coordination (for reviews and discussion, see Fusaroli, Konvalinka, & Wallot, in press; Riley & Shockley, in press), but also to quantify cognitive, neuro-, and motor dynamics, such as the learning of mathematical rules (Stephen, Boncoddo, Magnuson, & Dixon, 2009), the onset of epileptic seizures (Thomasson, Webber, & Zbilut, 2002), or the development of motor-language coordination (Abney, Warlaumont, Haussmann, Ross, & Wallot, 2013). Furthermore, RQA has been applied to response times of naming tasks (Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012) and reading times in text reading (O'Brien et al., 2013; Wallot et al., 2013), where recurrence measures of the stability of the time series have been shown to be indicative of reading skill.

The stability of reading times as understood by the measure %Determinism is linked to text comprehension through the attractor concept. If the behavior of a system has an attractor quality, the system displays ordered behavior in time. It does not bounce randomly through the phase space, but is rather confined to some kind of subsection of the phase space, thereby expressing some degree of repetitiveness, that can be captured by the measure of %Determinism. Successful text reading, that is, reading with comprehension, means that a reader progresses through a text with a growing understanding for its content, topics, and themes. This understanding, in turn, delimits the options for comprehending newly read words and sentences, as they are not isolated, but understood in terms of a particular context (e.g., Brisard, Frisson, & Sandra, 2001; Gibbs, 1984). Hence, the reading of words in a connected text is "biased" toward a particular meaning. Unlike in randomized single-word reading tasks, skilled text readers are attracted to a certain semantic space. %Determinism captures the degree to which this overarching semantic attractor exerts control over the reading process. In this sense, an attractor for reading is described as neither a property of the reader's behavior itself nor a property of the text itself, but instead a structured interaction between the two. Even though it might vary—during the course of reading, or between different readers—what information in a text is utilized, or how it is utilized, skilled readers will utilize *some* structure provided by a text. Hence, reading with comprehension should show a kind of global stability, despite trial-by-trial variations or interindividual differences in the reading process.

Relation of the Complexity Measures to Each Other and to the Reading Process

As these measures are fairly new in psychological research and as there is little precedent for their use in reading research, we summarize the most important points before moving on to introduce the reading study.

First, what is common across all three of these measures is that they capture *global* as opposed to *local* aspects of the reading process: Many psychological studies of reading ask how local aspects of a trial (such as the word frequency of a word in a lexical decision task) specifically affect the response of that particular trial (such as the time that it takes a participant to press a button in response to that word). The complexity measures described here offer a kind of complementary perspective to that, asking how the process of reading is systematically coordinated across many trials. Hence, we are interested in how reading globally affects measures of the reading process and, in particular, how it affects measures of the reading process differently for better readers (i.e., fast reading speed and good comprehension) versus poorer readers (i.e., slow reading speed with low comprehension).

To understand these metrics in the context of reading, it is helpful to think of a "null model" for our observable of the reading process: As the current reading study, described below in more detail, features a self-paced reading task, where participants press a response key to reveal every new word of the text (Just, Car-

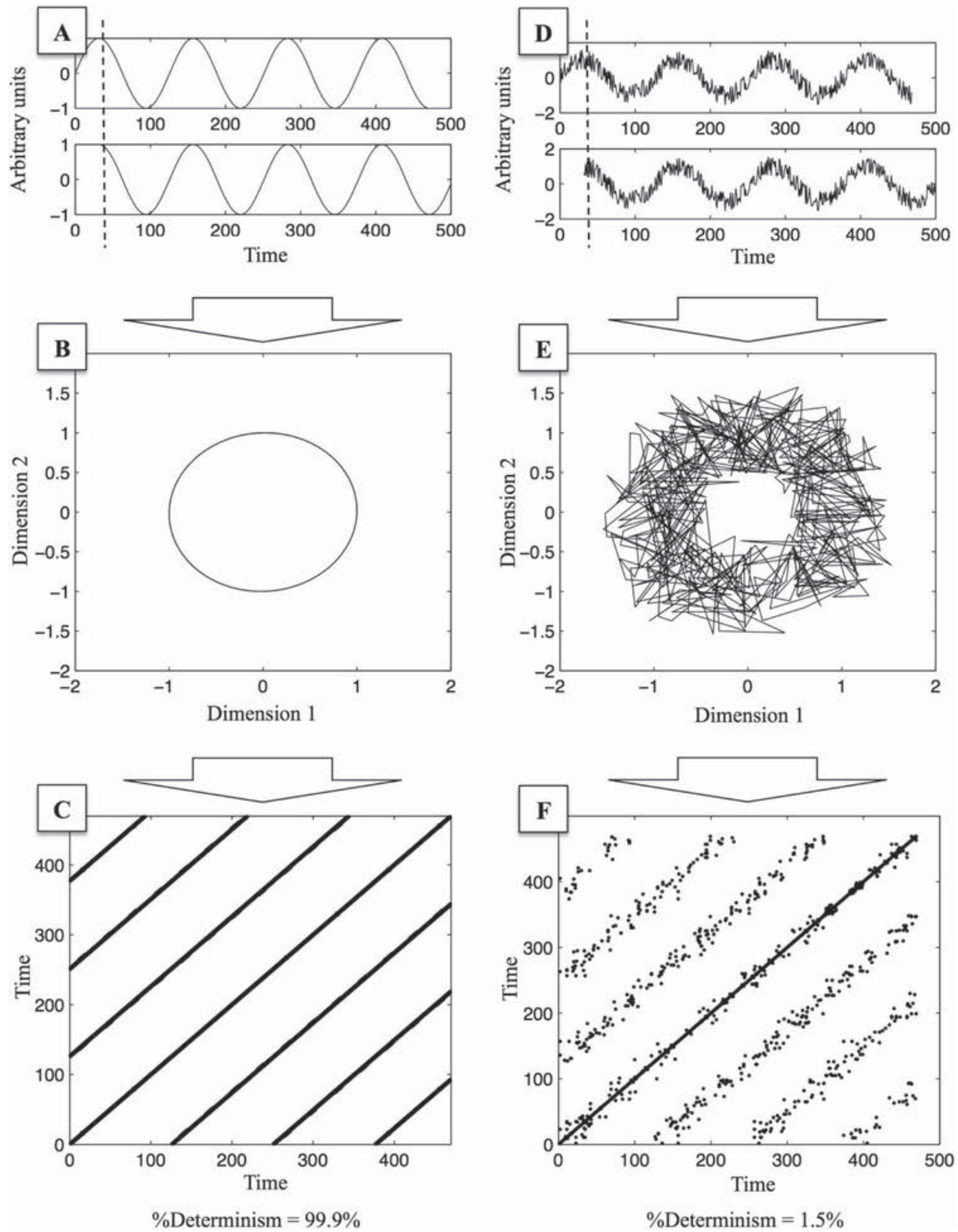

Figure 3. Illustration of phase space reconstruction from one-dimensional time series and the associated recurrence plots. (A) A sine wave with its time-delayed copy. (B) Phase space of the sine wave, reconstructed by plotting the original data series against its time-delayed copy. (C) Recurrence plot of a sine wave. All recurrent points of the time series form clear diagonal line patterns (dark lines in the plot). (D) The sine wave plus random noise with its time-delayed copy. (E) Phase space of the sine wave plus noise. (F) Recurrence plot of the sine wave plus noise. Recurrences of the time series (dots in the plot) cluster along diagonal lines in the recurrence plot, but are much more dispersed than without noise. The measure %Determinism is the sum of all dots in the plot that form connected diagonal lines divided by the sum of all dots in the plot. The more dots form the diagonal lines, the higher the %Determinism.

penter, & Woolley, 1982), our observable of the reading process is sequences of repeated keypresses. A basic null model for repeated keypresses in the absence of reading comes from tapping studies, where participants simply press a response key repeatedly without much information processing demands between keypresses. Here the global pattern of response times mainly reflects the endogenous (motor) activity that instantiates simple, repeated finger movements (Gilden, 2001). This activity is by no means random, but exhibits pink-noise-type long-range correlations that are well captured by monofractal scaling (see Figure 1). When keypresses are contingent on the need to process information between trials, such as when participants are provided with accuracy feedback between tapping trials, then endogenous activity is constraint by the exogenously provided information, leading to a decrease in long-range correlations. Furthermore, the more feedback is provided, the greater the decrease in monofractal scaling (Kuznetsov & Wallot, 2011). Hence, for monofractal scaling, we expect that the more attentively participants read, the more contingent their keypresses during self-paced reading will be on the information provided in the text, and the more reading times will depart from pink toward white noise.

However, information also structures keypresses. For the case of text reading, this can happen in many ways. For example, at the paragraph level, sentences are ordered in a particular way to systematically construct concepts that are not encapsulated within a single sentence, but transcend many sentences (e.g., Altmann, Cristadoro, & Esposti, 2012). At the sentence level, syntactic constructions mandate a particular ordering of words, so the integration of the different information pieces is not arbitrarily possible at any time, but preferably happens at the end of sentences, for example, resulting in wrap-up effects (e.g., Just & Carpenter, 1980). At the word level, pieces of information that characterize one word, such as its part of speech (e.g., Abrams & Rodriguez, 2005), semantic content (e.g., Perea & Rosa, 2002) or co-occurrence (e.g., Landauer & Dumais, 1997), prepare reading of adjacent words, resulting in local carryover effects. Insofar as readers' keypressing behavior during reading is contingent on the information in a text, such reoccurring structures in texts should be reflected in reading times, resulting in a greater sequential order of reading times that is captured by %Determinism.

An important note is in order here, as the predictions for monofractal scaling (= less long-range correlation with better comprehension) and %Determinism (= more sequential repetitiveness with better comprehension) might seem to contradict each other. The answer to this apparent contradiction lies in the way each measure captures temporal structure: Long-range correlations of the monofractal type capture a homogeneous, self-similar kind of activity, where reading of a text would unfold in the same way across words within sentences and words between sentence boundaries, and across words, sentences, and paragraphs alike. However, reading is more heterogeneous than that, showing systematic breaks and qualitative changes in reading activity, for example, at the sentence boundary (Just & Carpenter, 1980), but also occasional changes in reading activity, such as rereading of a sentence (Rayner et al., 2006) or long pauses during reading (O'Brien et al., 2013) that intersperse an otherwise well-structured process.

This leads us to the third measure of global reading activity that we employ: multifractal scaling. Multifractal scaling captures reorganizations of the reading process, such as the rereading of a sentence that was not properly understood, that are, on the one hand, a coordinated, systematic part of the reading process but, on the other hand, mark a strong departure from the "standard mode" of reading where readers give each line of the text just one pass that is organized along the line of text regularities, such as syntactic, semantic, and lexical features. One could say that multifractal structure distinguishes unsystematic "outlier" activity from reading activity that is an outlier in the sense that it is singular or rarely occurring, but a natural and systematic part of the reading process. Here one has to consider the functional relation between the standard mode of reading (such as captured by %Determinism) and the rarely occurring systematic departures from that mode: Understanding of a text develops through the text, where later parts of a text are understood based on what has been read previously. If this is done skillfully by a reader throughout the text, we expect that the reader's behavior (such as measured by word reading times) follows the salient structural aspects of the text (indexed by low monofractal scaling and high %Determinism). However, if earlier parts of the text are not well understood, and comprehension of later parts is dependent on that information, then there are basically two alternatives of our conception of the reading process. Either the reader starts a temporary but systematic departure from the reading mode. This could be, for example, rereading of an earlier passage or pausing to think about what was being read. This will ensure that the reader can enter the standard mode of reading after this temporal departure again, leading again to well-structured reading behavior—an effect that has been observed in mathematical rule-learning tasks (Stephen et al., 2009). Alternatively, if the reader does not resolve central ambiguities about the text, then the standard mode will be severely compromised, reading will become more difficult, and the reader will continue to depart from a well-structured reading process (e.g., Rayner, 1986). This again highlights our initial concern about the role of reading speed, where a certain level of reading speed can be reflective of both a time investment (careful reading, rereading, etc.) and severely compromised reading activity (problems with decoding, prolonged but unresolved uncertainty, etc.), but with very different consequences for the quality and outcome of the reading process.

Admittedly, much of the outlined logic regarding the three complexity measures stems from reasoning about what kind of structure they pick up in time series, or from what has been learned in other areas of psychology, and there is little empirical precedent for reading research. However, prior work of ours has at least shown the utility of monofractal and %Determinism measures to distinguish between text difficulty (Wallot, 2011a) and different stages of reading development (O'Brien et al., 2013) that fit with the outlined logic, respectively. By employing these three metrics, we thus examine the features of external constraints (monofractal), structuredness and stability (%Determinism), and adaptability (multifractal) on the reading process across a range of individuals' skill level and across oral and silent reading modes.

The Reading Study: Reading Times and Comprehension

In the present study, we reanalyzed the data from O'Brien, et al. (2013), a self-paced reading task resulting in reading time series data from children (second, fourth, and sixth graders) and adults (undergraduate students). All participants read the same text—an appropriate, moderately challenging text for second graders. In this self-paced reading task, readers pressed a response key to reveal every new word in the text, resulting in a series of word reading times for each participant. Half of the participants read the text orally, the other half silently. After reading the story, participants answered a multiple-choice questionnaire about the content of the text to assess their comprehension.

In the original study, O'Brien et al. (2013) showed that %Determinism of reading times increased with reader age, indicating that it captures developmental differences in reading fluency. No developmental trend was observed for monofractal scaling of reading times (the measure of multifractality was not employed). Comprehension scores were collected together with reading time data, but were not analyzed further.

Hence, in the present study, we pursue the relationship between aspects of reading fluency as captured by monofractal scaling, multifractal scaling, and %Determinism and their relation to text comprehension. As we are interested in the relation between reading process measures and reading outcome measures, the time-ordered reading time data are subjected to fractal and recurrence analyses. The results from these analyses are then correlated with the outcome measures of comprehension scores and overall reading speed (words per minute [WPM]). We also computed a comprehension ratio score: comprehension divided by reading time, as readers might "trade in" reading speed for gains in comprehension.

The main theoretical dimension of the dynamics of reading pertains to the adaptive fluency of the reading process, and how the reading outcome, measured by comprehension and speed, comes about as a result of this process. Given our theoretical outline of the meaning of complexity measures in the context of reading and text comprehension, we can summarize our hypotheses as follows: It has been argued that monofractal scaling represents the extent to which cognitive activity is driven by internal sources or external constraints. In the case of reading, it is commonly assumed that the ongoing informational input provided by the text strongly drives the reading process; hence, a reading process that is systematically guided by text structure should be an index of skilled reading, reflected in a negative relationship between the strength of monofractal scaling in reading times and comprehension scores. Multifractal structure in reading times is an index of cognitive reorganization, reflecting changes or developments in the understanding of the text's content. Hence, the presence of multifractal structure in reading times should correlate positively with text comprehension. Finally, if high comprehension acts as an attractor, constraining the ongoing reading activity within the context of what has already been read, then %Determinism, an index of attractor strength, should correlate positively with comprehension. In line with prior research findings, it is furthermore expected that reading speed and comprehension are positively correlated with each other as well.

Method

Participants

Sixty-two children between 6 and 13 years of age were recruited from urban elementary schools in the United States serving predominantly middle-class families of mixed racial background. Eight children (seven of whom were in the silent reading condition) were tested but excluded from the final sample because they did not advance far enough through the story and thus did not meet the inclusion criterion of having produced a data series of more than 900 data points. This criterion was adopted as a trade-off, considering that it would include those who had read at least five out of the six pages of the story, and that 900 data points are still sufficient to estimate reliable fractal scaling exponents (Holden, 2005). The final sample (26 boys and 28 girls) consisted of 19 second graders (M = 7.8 years, SD = 4.9 months), 20 fourth graders (M = 10.0 years, SD = 5.7 months), and 15 sixth graders (M = 12.0 years, SD = 3.9 months), as well as 17 adults (M = 24.5 years; six men, 11 women) who functioned as a comparison group. All participants were native speakers of English, and all had normal or corrected-to-normal vision. Also, all participants reported that they were unfamiliar with the story they read for this study. Participants were randomly assigned to either the silent reading condition (n = 35) or the read-aloud condition (n = 36).

Stimuli and Apparatus

Children read the story "Alien and Possum: Hanging Out" by Tony Johnston (2003). The text describes two friends, Alien and Possum, and the various things they do over several days (such as having conversations about their lives, celebrating birthdays, and climbing trees). The difficulty of the text was rated to be appropriate for somewhat advanced second graders (ATOS book level of 2.5), with the content being rated as interesting (4.5 out of 5) by readers ("Book Details," n.d.). The original text consisted of 1,201 words, but it was carefully trimmed down to 1,099 words to reduce reading time for the children, while at the same time satisfying the lower bound of data points necessary to conduct the analyses (see Holden, 2005; Wallot, O'Brien, & Van Orden, 2012). Besides, none of the illustrations in the original book were presented with the text. The text was displayed on a conventional 15-in. (38.10-cm) MacBook, with a custom script (MATLAB Psychophysics Toolbox; Brainard, 1997) running the text presentation software. Participants advanced through the text in a self-paced manner by pressing the space bar to reveal every new word of the text, with text accumulating on the screen word by word. The time between keypresses was taken to estimate the reading time of each word unit.

Procedure

Children were tested individually in a quiet area at their school, and adults were tested in a university laboratory room. A sample sentence was presented first, before commencing the story. Participants were instructed to read at their own pace for understanding, and that they would be given comprehension questions after they had completed the story. Participants in the silent reading condition were instructed to "read the words as they come up silently, to yourself." Participants in the read-aloud condition were instructed to "say the words as they come up—so you are reading out loud." Comprehension was assessed with six multiple-choice questions that included literal questions about events taken from different points in the story. The comprehension questions and multiple-choice answers were read to the participants as they read along, and then they gave their response. Each participant's session took between 10 and 20 min, and children were afterward rewarded with a little gift.

Data Analysis

In a first step, extreme response times of 10 s or longer were removed (eliminating 0.1% of all data points). Then the complexity metrics were computed. The measure %Determinism was obtained from recurrence computed with the Commandline Recurrence Plots software (Marwan, 2006). The RQA parameters were estimated following recommendations by Webber and Zbilut

(2005). Fractal scaling exponents were quantified with detrended fluctuation analysis (Peng, Havlin, Stanley, & Goldberger, 1995) following recommendation by Holden (2005). Multifractal spectra were quantified with multifractal detrended fluctuation analysis (Kantelhardt et al., 2002) following recommendations by Ihlen and Vereijken (2010). Details about the analysis techniques can be found in the Appendix.

Pearson correlations were used to assess the relationship between the different metrics of reading times and reading speed, raw comprehension scores, and ratio comprehension scores where comprehension was divided by reading time. Finally, stepwise regression analysis was used in order to select the best predictors of comprehension from all variables of the reading process that were computed.

Results

Reading Rate and Comprehension

As reported by O'Brien et al. (2013), mean reading rates (WPM) differed between grades; that is, speed increased with age as expected, $F(3, 66) = 19.44$, $p < .001$. Furthermore, we observed a Grade × Reading Mode interaction, indicating that reading rate differed across modes only for Grade 2 and adult groups, $F(3, 66) = 6.63$, $p < .001$, as planned contrasts of oral versus silent reading only reached significance for Grade 2 and adult groups (ps $< .003$; see Table 1). Comprehension (correct responses) also increased with age, $F(3, 66) = 6.72$, $p < .001$. With respect to the reading mode conditions (oral vs. silent), comprehension was equivalent across modes, $F(1, 66) = 1.26$, $p = .266$.

Reading Time Statistics

When looking at developmental differences in the structure of reading times, O'Brien et al. (2013) found that WPM and %Determinism, but not monofractal scaling, differed significantly across age groups. In order to rule out that the effects are primarily driven by a general maturational factor, we ran correlational analyses and controlled for reader age. As can be seen in Table 2, when age is controlled for, reading speed and comprehension are significantly correlated only for oral reading and not during silent reading. Further, %Determinism for both oral and silent reading shows a significant and stronger relation to comprehension than does reading speed. On the other hand, monofractal structure does not show any significant relationships with comprehension, and the expected positive relation between multifractality and comprehension is only observed during silent reading. All of these effects are corroborated when the ratio score is used as a measure for comprehension. As the ratio score equals correctly answered comprehension questions divided by overall reading time, the correlation between reading speed and the ratio score is trivial. For the interested reader, we also present the correlations between age, speed, comprehension, and the complexity measures in the Appendix.

Table 1
Averages and Standard Deviations of Reading Rate and Comprehension Scores for Oral and Silent Reading by Grade

	Reading rate (WPM)		Comprehension (%)	
Grade	Oral	Silent	Oral	Silent
2	43.0 (10.7)	92.4 (35.8)	66.6 (28.5)	57.3 (31.2)
4	78.2 (26.4)	78.7 (22.7)	85.4 (17.6)	88.1 (23.0)
6	90.1 (21.5)	88.8 (22.5)	85.6 (17.7)	91.5 (16.5)
Adults	112.1 (28.3)	174.3 (34.3)	100 (0.0)	98.1 (5.7)

Note. Reading rate is reported as average words per minute (WPM), and comprehension is reported as percent correct on average for the groups. Four answer options per multiple-choice question for each comprehension item make for a chance performance value of 25% correct. Standard deviations are in parentheses.

Table 2
Correlations Between Reading Time Statistics and Comprehension Measures When Controlling for Age

Variable	1	2	3	4	5	6
Oral reading (interkeypress intervals) with age parceled out						
1. WPM	—	.332*	.798**	.602**	.045	.395*
2. Comp.		—	.664**	.566**	.004	.018
3. Ratio			—	.652**	−.060	.174
4. DET				—	.250	.265
5. DFA					—	.289
6. MFDFA						—
Silent reading (interkeypress intervals) with age parceled out						
1. WPM	—	.278	.761**	.698**	−.353*	.402*
2. Comp.		—	.714**	.397*	.019	.348*
3. Ratio			—	.608**	−.183	.398*
4. DET				—	.000	.450**
5. DFA					—	−.210
6. MFDFA						—

Note. Correlation coefficients without asterisks are not statistically significant. WPM = words per minute; comp. = comprehension scores; ratio = ratio comprehension scores (comprehension scores are divided by reading time); DET = %Determinism; DFA = monofractal scaling estimated by detrended fluctuation analysis; MFDFA = multifractality estimated by multifractal detrended fluctuation analysis.
* $p < .05$. ** $p < .01$.

Regression of Reading Outcome Onto Reading Process Variables

To assess which aspects of the reading process are unique predictors for comprehension, we used a stepwise regression procedure where predictors were successively added to the model (in the order WPM, %Determinism, monofractal scaling, multifractal scaling). During each step of the procedure, a predictor was first added to the model. Then all predictors were evaluated as to whether they still contributed unique predictive power to the model (at $\alpha = .05$), and those that did not (anymore) were removed from the model.

As can be seen in Tables 3 and 4, %Determinism of reading times came out as the single and best predictor for comprehension during oral and silent reading. For the ratio score, %Determinism of reading times was also the best predictor. For the oral reading condition, monofractal fluctuations of reading times also added unique predictive power.

Table 3
Stepwise Regression of Comprehension Scores

%Determinism	β	*t*	*p*
Oral reading	.566	4.07	$<.000$
Silent reading	.397	2.60	.014

Note. The model for oral reading explains $R^2 = .321$ of the variance in comprehension scores, $F(1, 34) = 16.07$, $p < .001$. The model for silent reading explains $R^2 = .158$ of the variance in comprehension scores, $F(1, 35) = 6.56$, $p = .015$.

Word Frequency and Length

The distribution of lexical properties in texts has been shown to exhibit complex scaling patterns (Ebeling & Poeschel, 1994; Montemurro & Pury, 2002). Hence, the observed complexity in reading times might have its origin in lexical text properties. To investigate whether this is the case, we apply the monofractal, multifractal, and %Determinism measure to the vector of word lengths and word frequencies of the text used in this study. Figure 4 shows these vectors together with four examples of reading time series from second graders and adult readers.

Visual inspection of the data series suggests some structural differences between reading times and lexical properties. Compared to the word length and frequency vector, reading times seem to be composed of a more extreme variation between pockets of relative uniformity (i.e., many equally fast reading times in a row) and pockets of extreme fluctuations. To quantify this first impression, Table 5 presents the monofractal, multifractal, and %Determinism properties of the word length and word frequency vectors, together with the values for these measures for silent and oral reading times. Word frequencies were taken from the norms of the English Lexicon Project (Balota et al., 2007) and were logarithmized. As can be seen, word length and frequency lie outside the 95% confidence interval of the reading time data on all three complexity metrics, compared to reading time data (see Table 6). Furthermore, correlations between word length and frequency with reading times are relatively low (see Table 6). Given that word length and word frequency are highly correlated with each other ($r = -.724$), together they explain only 1.3% (oral reading) and 0.7% (silent reading) of the variance in reading times, respectively. In sum, this suggests that lexical properties are at least not in a simple, straightforward manner the source of complexity in reading times during text reading.

Discussion

In search for new characterizations of fluency of the reading process that do not rely solely on speed or duration, we hypothesized that good comprehension during reading is reflected by reading time patterns with high temporal stability, but also interspersed with adaptive fluctuations. We presented three metrics that measure the external constraint, adaptability, and stability of the reading process: the influence of external information on time series measurements (monofractal scaling), the occurrence of reorganizations during the reading process (multifractal scaling), and the degree of temporal structuredness of reading times (%Determinism).

The %Determinism measure showed consistent effects across oral and silent reading, predicting absolute comprehension scores, as well as comprehension ratio scores, when comprehension scores were normalized by reading time. Against our expectations, monofractal scaling did not show significant relations with comprehension. We only observed a moderate correlation with reading speed during silent reading. Multifractal scaling, on the other hand, did show the expected relations to the comprehension measures, but only for silent reading. There was also a moderate correlation with reading speed in both the oral and silent reading conditions. Reading speed itself, however, showed only a moderate correlation to comprehension during oral reading (faster reading goes along with better comprehension), while this was not the case for silent reading.

There may be several reasons for the lack of relations between monofractal scaling and comprehension. First of all, these metrics have not been much employed in reading research, and hence our proposal regarding the conceptual role of this measure might be flawed or misguided. For example, in keeping with the early literatures on monofractal scaling (e.g., Wijnants et al., 2009), we previously predicted that a Hurst exponent closer to pink noise would be related to more fluent reading, following findings of fractal scaling for more coordinated behavior (O'Brien et al., 2013). In a more recent conceptualization (Van Orden et al., 2011), on the other hand, the prediction was opposite, based on the model that fluent reading, in terms of automatic word recognition, would be externally driven by the text and thus appear closer to white noise (compared with pink noise). Therefore, before discarding the potential utility of monofractal scaling as an interesting measure for reading, two alternatives could be considered.

First, monofractal scaling might capture how far the informational context impinges on the endogenous fluctuations of the time series measure of reading. As we laid out, simple repeated keypresses that are not contingent on informational processing of stimuli exhibit strong monofractal long-range correlations (pink noise). Pink noise seems to be a basic feature of unconstraint motor-physiological processes. Besides keypresses, it is found in several other observables that are used to measure reading, such as eye movement fluctuations of the resting eye (Coey, Wallot, Richardson, & Van Orden, 2012) and resting-state functional magnetic resonance imaging (fMRI; He, 2011). Pink-noise-type long-range correlations seem to capture general endogenous motor and physiological processes.

In a task context, where the motor-physiological processes behind these observables have to sustain the performance of a specific function, usually decreases of long-range correlations toward white noise are observed, such as in the keypress fluctuations

Table 4
Stepwise Regression of Ratio Comprehension Scores

%Determinism	β	*t*	*p*
Oral reading	.652	5.09	$<.000$
Silent reading	.608	4.68	$<.000$

Note. The model for oral reading explains $R^2 = .425$ of the variance in comprehension ratio scores, $F(1, 34) = 25.17$, $p < .001$. The model for silent reading explains $R^2 = .379$ of the variance in comprehension ratio scores, $F(1, 35) = 21.32$, $p < .001$.

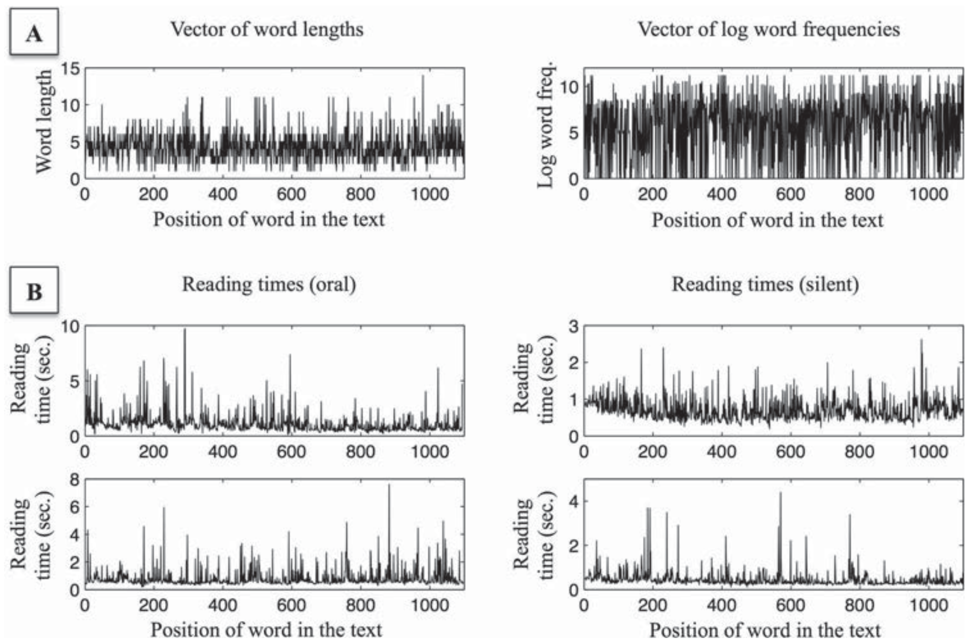

Figure 4. Vectors of lexical properties and example reading times. (A) Vectors of word length and log word frequency (freq.) for the text. (B) Example time series of oral and silent reading displaying a second grader's (top row) and an adult's (bottom row) reading time for each mode. Visual, lexical properties seem to differ form reading times in their degree of heterogeneity.

during self-paced reading (O'Brien et al., 2013), in eye movements during reading (Wallot, 2011b), and in the fMRI signal during reaction time performance (He, 2011). Hence, monofractal scaling may primarily measure the difference between "reading" the text by pressing a response key in accordance with the unfolding of the text and simply pressing the button repeatedly in the absence of information. This would stand in contrast to our proposal that monofractal scaling reflects how well participants utilize linguistic features of written language during reading. It might rather measure that a participant is reading, but not how well a participant does so. Accordingly, decreases in monofractal scaling might just distinguish between participants who are actually reading and those who press the button without paying attention.

Second, the developmental nature of our sample might have worked against the effect of monofractal fluctuations. It has been shown that simple button press performance develops from white noise in young children toward pink noise in adults (Kiefer et al., 2014), indicating improvement in the fluency of fine motor coordination. For the case of reading here, we predicted that the reading skill, as reflected in comprehension, should have the opposite effect on monofractal scaling in reading times, being closer to pink noise in the case of poor comprehension and closer to white noise in the case of good comprehension. As younger readers are expected to show a lower degree of fluency in motor coordination in button presses (i.e., white noise), but also poorer reading skill and comprehension (i.e., pink noise), these two effects might have canceled each other out.

Table 5
%Determinism, Monofractal, and Multifractal Characteristics of Lexical Text Properties and Reading Times

			Reading times			
			Oral		Silent	
Variable	Word length	Log word frequency	Average	95% CI	Average	95% CI
DET	0.521	0.700	0.917	[0.891, 0.942]	0.958	[0.940, 0.976]
DFA	0.542	0.515	0.570	[0.547, 0.594]	0.611	[0.579, 0.644]
MFDFA	0.076	0.040	0.300	[0.254, 0.346]	0.488	[0.423, 0.554]

Note. Monofractal, multifractal, and %Determinism (DET) values for word length and the logarithm of word frequency. Furthermore, the averages of monofractal, multifractal, and DET values for the whole sample of oral and silent readers, together with the associated 95% confidence intervals (CIs). DFA = monofractal scaling estimated by detrended fluctuation analysis; MFDFA = multifractality estimated by multifractal detrended fluctuation analysis.

Table 6
Average Correlation Coefficients for Word Length and Word Frequency With Reading Times

Variable	Oral		Silent	
	r	95% CI	*r*	95% CI
Word length	.099	[.081, .116]	.072	[.053, .091]
Log word frequency	−.134	[−.114, −.153]	−.101	[−.081, −.121]

Note. Average correlations between word length and the logarithm of word frequency with reading times during oral and silent reading, together with the associated 95% confidence intervals (CIs). The effect of word length and word frequency on reading times is weak, but reliable at the 95% CI. Using the nonlogarithmized word frequency or nonlogarithmized reading times led to decrease of the effect size, while further trimming of the reading times (e.g., discarding all reading times greater than 2.5 s) did not improve the correlations.

That multifractal scaling showed the expected relations to comprehension in the silent reading condition is encouraging, since silent reading performance has been much harder to reliably quantify compared to oral reading (Price, Meisinger, D'Mello, & Louwerse, 2012). The absence of effects in oral reading might have something to do with the somewhat different nature of button presses during oral reading. Traditionally, reading times during oral reading are measured from stimulus onset to the articulatory onset (Kessler, Treiman, & Mullennix, 2002). In our case, the whole articulation phase of a word was part of the recorded reading times, which might have resulted in a decreased sensitivity of the oral reading condition with regard to the multifractal scaling measure. The additional time provided by the sounding out of a word might have dampened out some of the fluctuations due to change in text comprehension that are evident in the silent reading times, which are putatively more reflective of the "raw content" of the text.

Our findings concerning the relation between reading speed and straight comprehension scores mirror the somewhat mixed effects that have been reported in the literature. The overall correlations in the present study are rather moderate, and the effect for silent reading did not reach significance. The asymmetry of oral and silent reading in this regard might highlight that the compensatory role of reading speed for comprehension plays a bigger role in the faster paced silent reading, compared to reading out loud, obscuring the overall relationship. This seems to be corroborated by differences observed in the stepwise regressions for the raw comprehension scores compared to the ratio score, where comprehension was normalized by reading time.

When we used stepwise regression to select the best predictors for comprehension, the measure of %Determinism was retained in the model in favor of speed and yielded significant effects across both reading modes. When comprehension was normalized by reading time, the explanatory power of the %Determinism measure increased—especially for silent reading. That is, the relationship between predictors derived from the reading process and the overall outcome measure shows strong and comparable effect sizes across both reading modes. This is important to highlight, because the bulk of literature on reading fluency and comprehension is concerned with oral reading (Share, 2008; Kim, Wagner, & Foster, 2011), and reading metrics for silent reading are currently limited (e.g., Fuchs et al., 2001; Price et al., 2012). Taking speed as an outcome rather than a process measure disambiguates the recorded comprehension scores. The similar relations of the ratio scores to %Determinism, that is, the stability of the reading times, suggest a common core that characterizes the reading process in both tasks.

However, these commonalities do not seem to stem from the lexical properties of the text. Even though corpus analyses have shown that variables such as word frequency show complex fractal distributions in texts (Ebeling & Poeschel, 1994; Montemurro & Pury, 2002), an analysis of the lexical structure of the text used in this study suggested that it differed substantially from the patterns observed in reading times. Furthermore, correlations between reading times and word frequency were generally very low, corroborating earlier research that found that lexical variables do not play much of a role in text reading (Wallot et al., 2013). Naturally, this casts up the question as to what the salient structure is, which readers pick up on in text reading. As our study did not contain variations of different texts or text features, we cannot at present answer this question. However, more abstract, entropy-related text metrics might prove to be a viable next step for further investigations into what the salient textual structures are that readers utilize during reading (Grenzel & Charniak, 2002; Montemurro & Zanette, 2011). These metrics capture not only general lexical but also idiosyncratic contextual variations in text, which might be a key difference between reading situations that feature isolated words or sentences compared to connected texts.

Hence, even though more work is needed in order to clarify the role of fractal and recurrence variables in reading and text comprehension and their grounding in a psychological theory of the reading process, the results of the present study point to the promising role of complexity metrics for the investigation of complex cognitive tasks such as reading. Apart from their utility as predictors of reading performance outcomes, these metrics might furthermore provide the missing link that binds together features of texts and cognitive abilities across words, trials, and timescales.

Conclusion

In this study, we examined how complexity metrics of reading times that measure structuredness and adaption of the reading process can be used to predict aspects of reading skill, such as text comprehension. This investigation was motivated by conceptual and empirical problems with reading speed as a measure of the reading process. Conceptually, the role of reading speed is ambiguous, because it is used both as an outcome measure of reading and as a measure for the reading process. And as a process measure it is problematic, because it conflates two roles for reading speed in text reading, namely, speed as an index of fluency and speed as a compensatory resource. Empirically, reading speed is problematic, because it is not a reliable predictor of text comprehension, the hallmark of skilled reading.

Our results corroborate that reading speed is not a reliable predictor of text comprehension across silent and oral reading. Instead, the degree of temporal structure in reading times (%Determinism) turned out to be a good predictor of comprehension across conditions. Furthermore, the predictive power of structure in reading times increased when comprehension scores were normalized by speed, highlighting the multiple roles that speed can play in reading. The results for other complexity metrics (mono- and multifractal scaling) were less clear, showing mainly effects for silent reading, which might be due to the fact that self-paced reading, as we employed it in the current study, is not a good way to measure oral reading.

Finally, analyses of word length and word frequency yielded only minimal effects of lexical properties on reading times in text reading, explaining only around 1% of the variance in reading times. This implies that the complex structure of reading times is not a simple function of the distribution of word properties in a text. Also, it suggests that the textual aspects that structure reading times during text reading do not reside on the level of individual words. Differently from reading isolated words or sentences, reading of connected text might not be so much a one-way relation from linguistic properties to cognitive processes, but rather emerges as an interaction between reading process and text structure. A deeper investigation of the matter might warrant further research that focuses on reading of complex text stimuli, and it might need measures of the reading process that can be applied to complex reading behavior—such as the ones we presented here.

References

Abney, D. H., Wallot, S., Warlaumont, A. S., Haussmann, A., & Ross, J. M. (2013). *A case study of infant motor and vocal development using non-linear methods.* Manuscript submitted for publication.

Abrams, L., & Rodriguez, E. L. (2005). Syntactic class influences phonological priming of tip-of-the-tongue resolution. *Psychonomic Bulletin & Review, 12,* 1018–1023. doi:10.3758/BF03206437

Altmann, E. G., Cristadoro, G., & Esposti, M. D. (2012). On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 11582–11587. doi:10.1073/pnas.1117723109

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258–290. doi:10.1037/h0055756

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language, 49,* 183–200. doi:10.1016/S0749-596X(03)00027-5

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Berninger, V. W., Abbott, R. D., Trivedi, P., Olson, E., Gould, L., Hiramatsu, S., . . . Westhaggen, S. Y. (2010). Applying the multiple dimensions of reading fluency to assessment and instruction. *Journal of Psychoeducational Assessment, 28,* 3–18. doi:10.1177/0734282909336083

Book details: *Alien & Possum: Hanging out*. (n.d.). Retrieved from http://www.arbookfind.com/bookdetail.aspx?q=59508&l=EN&slid=194455065

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Brisard, F., Frisson, S., & Sandra, D. (2001). Processing unfamiliar metaphors in a self-paced reading task. *Metaphor and Symbol, 16,* 87–108. doi:10.1207/S15327868MS1601&2_7

Caccia, D. C., Percival, D., Cannon, M. J., Raymond, G., & Bassingthwaighte, J. B. (1997). Analyzing exact fractal time series: Evaluating dispersional analysis and rescaled range methods. *Physica A, 246,* 609–632. doi:10.1016/S0378-4371(97)00363-4

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading, 36,* 84–95.

Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long memory processes ($1/f^{\alpha}$ Type) in human coordination. *Physical Review Letters, 79,* 4501–4504. doi:10.1103/PhysRevLett.79.4501

Coey, C. A., Wallot, S., Richardson, M. J., & Van Orden, G. (2012). On the structure of measurement noise in eye-tracking. *Journal of Eye Movement Research, 5,* 1–10.

Coltheart, M., Rastle, K., Conrad, P., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. doi:10.1037/0033-295X.108.1.204

Donald, M. (2007). The slow process: A hypothetical cognitive adaption for distributed cognitive networks. *Journal of Physiology—Paris, 101,* 214–222. doi:10.1016/j.jphysparis.2007.11.006

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22,* 389–406. doi:10.2307/747699

Ebeling, W., & Poeschel, T. (1994). Entropy and long-range correlations in literary English. *Europhysics Letters, 26,* 241. doi:10.1209/0295-5075/26/4/001

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. doi:10.1037/0033-295X.112.4.777

Frazier, L., Clifton, C., Rayner, K., Deevy, P., Koh, S., & Bader, M. (2005). Interface problems: Structural constraints on interpretation? *Journal of Psycholinguistic Research, 34,* 201–231. doi:10.1007/s10936-005-3638-1

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256. doi:10.1207/S1532799XSSR0503_3

Fusaroli, R., Konvalinka, I., & Wallot, S. (in press). Analyzing social interactions: The promises and challenges of using cross recurrence quantification analysis. *Springer Proceedings in Mathematics & Statistics.*

Gibbs, R. W. (1984). Literal meaning and psychological theory. *Cognitive Science, 8,* 275–304. doi:10.1207/s15516709cog0803_4

Gilden, D. L. (2001). Cognitive emission of $1/f$ noise. *Psychological Review, 108,* 33–56. doi:10.1037/0033-295X.108.1.33

Goldberger, A. L., Amaral, L. A. N., Hausdorff, J. M., Ivanov, P. Ch., Peng, C.-K., & Stanley, H. E. (2002). Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 2466–2472. doi:10.1073/pnas.012579499

Good, R., III, & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565. doi:10.1037/0033-295X.103.3.518

Grenzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In P. Isabelle (Ed.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 199–206). Stroudsburg, PA: Association for Computational Linguistics. doi:10.3115/1073083.1073117

He, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *Journal of Neuroscience, 31,* 13786–13795. doi:10.1523/JNEUROSCI.2111-11.2011

Hintze, J. M., Callahan, J. E., Matthews, W. J., Williams, S. A., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review, 31,* 540–553.

Holden, J. G. (2005). Gauging the fractal dimension of response times from cognitive tasks. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 267–318). Retrieved from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

Holden, J. G., Choi, I., Amazeen, P. G., & Van Orden, G. (2011). Fractal $1/f$ dynamics suggest entanglement of measurement and human performance. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 935–948. doi:10.1037/a0020991

Holden, J. G., Riley, M. A., Gao, J., & Torre, K. (2014). *Fractal analyses: Statistical and methodological innovations and best practices*. Retrieved from http://www.frontiersin.org/Fractal_Physiology/researchtopics/Fractal_Analyses_Statistical_a/555

Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review, 116,* 318–342. doi:10.1037/a0014849

Hosp, M. K., & Fuchs, L. S. (2005). CBM as an indicator of decoding, word reading, and comprehension: Do relations change with grade? *School Psychology Review, 34,* 9–26.

Ihlen, E. A. F. (2012). Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology, 3,* 141. doi:10.3389/fphys.2012.00141

Ihlen, E. A. F., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/*f* fluctuation. *Journal of Experimental Psychology: General, 139,* 436–463. doi:10.1037/a0019098

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95,* 719–729. doi:10.1037/0022-0663.95.4.719

Johnston, T. (2003). *Alien and possum: Hanging out*. New York, NY: Aladdin.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354. doi:10.1037/0033-295X.87.4.329

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes and in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228–238. doi:10.1037/0096-3445.111.2.228

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A, 316,* 87–114. doi:10.1016/S0378-4371(02)01383-3

Kello, C. T., Brown, G. D. A., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences, 14,* 223–232. doi:10.1016/j.tics.2010.02.005

Kelty-Stephen, D. G., Palatinus, K., Saltzman, E., & Dixon, J. A. (2013). A tutorial on multifractality, cascades, and interactivity for empirical time series in ecological science. *Ecological Psychology, 25,* 1–62. doi:10.1080/10407413.2013.753804

Kennel, M. B., Brown, R., & Abarbanel, H. D. I. (1992). Determining embedding dimension of phase-space reconstruction using a geometrical construction. *Physical Review A, 45,* 3403–3411. doi:10.1103/PhysRevA.45.3403

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47,* 145–171. doi:10.1006/jmla.2001.2835

Kiefer, A. W., Gresham, L. J., Wallot, S., Kloos, H., Riley, M. A., Shockley, K., & Van Orden, G. (2014). Development of coordination in time estimation. *Developmental Psychology, 50,* 393–401. doi:10.1037/a0033629

Kim, Y., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first-grade readers. *Scientific Studies of Reading, 15,* 338–362. doi:10.1080/10888438.2010.493964

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kloos, H., & Van Orden, G. (2010). Voluntary behavior in cognitive and motor tasks. *Mind & Matter, 8,* 19–43.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95,* 3–21. doi:10.1037/0022-0663.95.1.3

Kuznetsov, N. A., & Wallot, S. (2011). Effects of accuracy feedback on fractal characteristics of time estimation. *Frontiers in Neuroscience, 5,* 62. doi:10.3389/fnint.2011.00062

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323. doi:10.1016/0010-0285(74)90015-2

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240. doi:10.1037/0033-295X.104.2.211

LeVasseur, V. M., Macaruso, P., Palumbo, L. C., & Shankweiler, D. (2006). Syntactically cued text facilitates oral reading fluency in developing readers. *Applied Psycholinguistics, 27,* 423–445. doi:10.1007/s11145-007-9070-1

LeVasseur, V. M., Macaruso, P., & Shankweiler, D. (2008). Promotion gains in reading fluency: A comparison of three approaches. *Reading and Writing, 21,* 205–230. doi:10.1017/S0142716406060346

Mandelbrot, B. (1997). *Fractals and scaling in finance*. New York, NY: Springer. doi:10.1007/978-1-4757-2763-0

Marwan, N. (2006). Commandline Recurrence Plots (Version 1.13z) [Computer software]. Retrieved from http://tocsy.pik-potsdam.de/commandline-rp.php

Marwan, N. (2014). Recurrence plots and cross recurrence plots. Retrieved from http://www.recurrence-plot.tk/

McNerney, M. W., Goodwin, K. A., & Radvansky, G. A. (2011). A novel study: A situation model analysis of reading times. *Discourse Processes, 48,* 453–474. doi:10.1080/0163853X.2011.582348

Montemurro, M. A., & Pury, P. A. (2002). Long-range fractal correlations in literary corpora. *Fractals, 10,* 451–461. doi:10.1142/S0218348X02001257

Montemurro, M. A., & Zanette, D. H. (2011). Universal entropy of word ordering across linguistic families. *PLoS ONE, 5,* e19875. doi:10.1371/journal.pone.0019875

O'Brien, B. A., Wallot, S., Haussmann, A., & Kloos, A. (2014). Using complexity metrics to assess silent reading fluency: A cross-sectional study comparing oral and silent reading. *Scientific Studies of Reading, 18,* 235–254. doi:10.1080/10888438.2013.862248.

Orsucci, F., Giuliani, A., Webber, C., Zbilut, J., Fonagy, P., & Mazza, M. (2006). Combinatorics and synchronization in natural semiotics. *Physica A, 361,* 665–676. doi:10.1016/j.physa.2005.06.044

Paulson, E. J. (2005). Viewing eye movements during reading through the lens of chaos theory: How reading is like the weather. *Reading Research Quarterly, 40,* 338–358. doi:10.1598/RRQ.40.3.3

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos, 5,* 82–87. doi:10.1063/1.166141

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research, 66,* 180–194. doi:10.1007/s00426-002-0086-5

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Price, K., Meisinger, E., D'Mello, S. K., & Louwerse, M. (2012). Silent reading fluency with underlining: A new method of assessment. *Psychology in the Schools, 49,* 606–618. doi:10.1002/pits.21613

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41,* 211–236. doi:10.1016/0022-0965(86)90037-8

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10,* 241–255. doi:10.1207/s1532799xssr1003_3

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–476. doi:10.1017/S0140525X03000104

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Neuroscience, 7,* 361. doi:10.3389/fnhum.2013.00361

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65,* 177–206. doi:10.1016/j.cogpsych.2012.03.001

Riley, M. A., & Shockley, K. (in press). Interpersonal couplings in human interaction. In C. L. Webber Jr. & N. Marwan (Eds.), *Recurrence quantification analysis: Theory and best practices.* Springer.

Riley, M. A., & Van Orden, G. C. (2005). *Tutorials in contemporary nonlinear methods for the behavioral sciences*. Retrieved from https://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology, 103,* 877–896. doi:10.1037/a0023731

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of over-reliance on an "outlier" orthography. *Psychological Bulletin, 134,* 584–615. doi:10.1037/0033-2909.134.4.584

Stephen, D. G., Anastas, J. R., & Dixon, J. A. (2012). Scaling in executive control reflects multiplicative multifractal cascades. *Frontiers in Physiology, 3*, 102. doi:10.3389/fphys.2012.00102

Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition, 37,* 1132–1149. doi:10.3758/MC.37.8.1132

Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics, 898,* 366–381. doi:10.1007/BFb0091924

Thomasson, N., Webber, C., & Zbilut, J. (2002). Application of recurrence quantification analysis to EEG signals. *International Journal of Computers and their Applications, 9,* 9–14.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Van Orden, G., Holden, J. G., Podgornik, M. N., & Aitchison, C. S. (1999). What swimming says about reading: Coordination, context, and homophone errors. *Ecological Psychology, 11,* 45–79. doi:10.1207/s15326969eco1101_2

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General, 132,* 331–350. doi:10.1037/0096-3445.132.3.331

Van Orden, G., Kloos, H., & Wallot, S. (2011). Living in the pink: Intentionality, wellbeing, and complexity (pp. 629–672). In C. A. Hooker (Ed.), *Philosophy of complex systems: Handbook of the philosophy of science*. Amsterdam, the Netherlands: Elsevier.

Wallot, S. (2011a). *The role of reading fluency, text difficulty and prior knowledge in complex reading tasks* (Doctoral dissertation). Retrieved from https://etd.ohiolink.edu/

Wallot, S. (2011b, May). *Understanding as a constraint on behavior: The dynamics of reading reveal a slow process in text comprehension*. Paper presented at the Eighth Conference of the Nordic Association for Semiotic Studies Aarhus, Denmark.

Wallot, S., Hollis, G., & van Rooij, M. M. J. W. (2013). Connected text reading and differences in text reading fluency in adult readers. *PLoS ONE, 8*, e71914. doi:10.1371/journal.pone.0071914

Wallot, S., O'Brien, B. A., & Van Orden, G. (2012). A tutorial introduction to fractal and recurrence analyses of reading (pp. 395–430). In G. Liben, G. Jarema, & C. Westbury (Eds.), *Methodological and analytical frontiers in lexical research*. Amsterdam, the Netherlands: Benjamins.

Wallot, S., & Van Orden, G. (2011). Toward a lifespan metric of reading fluency. *International Journal of Bifurcation and Chaos, 21*, 1173–1192. doi:10.1142/S0218127411028982

Webber, C. L., Jr., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology, 76,* 965–973.

Webber, C. L. Jr., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26–96). Retrieved from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests (GORT-4).* Austin, TX: Pro-Ed.

Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. F. A., & Van Orden, G. C. (2009). 1/*f* scaling in movement time changes with practice in precision aiming. *Nonlinear Dynamics, Psychology, and the Life Sciences, 13,* 79–98.

Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia, 62,* 100–119. doi:10.1007/s11881-012-0067-3

Zbilut, J. P., Giuliani, A., & Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A, 246,* 122–128. doi:10.1016/S0375-9601(98)00457-5

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 286–397. doi:10.1037/0278-7393.21.2.386

(*Appendix follows*)

Appendix

Analysis Techniques

Correlations With Age

Table A1 presents the correlation coefficients of reading speed, comprehension, and the complexity metrics with age, for silent and oral reading.

Introduction to Fractal and Recurrence Analysis

This last part of the Appendix aims at giving a brief, step-by-step explanation of the monofractal, multifractal, and recurrence analyses techniques. After the description of each analysis technique, we summarize the parameter settings that were used to generate the results for this article. As part of an Appendix, this overview will necessarily stay somewhat limited in its scope. Therefore, we also compiled a list of more extended introductions to each analysis technique and summaries for best practices for the interested reader at the end of the Appendix.

Detrended fluctuation analysis (DFA) for monofractal scaling. We used DFA to estimate the degree of long-range correlation present in each time series of reading times, also called monofractal scaling (Peng et al., 1995). As we outlined in the introduction, monofractal scaling in a time series can be quantified with DFA as a scaling relation between variation in a time series and length of a time series. One can also think of the analysis as providing a measure for how variation in a time series changes across different timescales. We use as an example a reading time series from a fourth grader during silent reading (see Figure A1, left panel) to illustrate the process.

Table A1
Correlations Between Age, Reading Time Statistics, and Comprehension Measures

Age (months)	WPM	Comp.	Ratio	DET	DFA	MFDFA
Oral reading	.569**	.428**	.680**	.441**	−.200	−.128
Silent reading	.621**	.401*	.796*	.405*	−.119	.016

Note. Correlations between age, reading speed comprehension, and complexity metrics. Correlation coefficients without asterisks are not statistically significant. WPM = words per minute; comp. = comprehension scores; ratio = ratio comprehension scores (comprehension scores are divided by reading time); DET = %Determinism; DFA = monofractal scaling estimated by detrended fluctuation analysis; MFDFA = multifractality estimated by multifractal detrended fluctuation analysis.
* $p < .05$. ** $p < .01$.

First, the type of the time series needs to be classified. If the time series *x* is noise (as opposed to a random walk), then it is first integrated (see Figure A1, right panel), yielding the integrated time series *y*. As we are interested in how variation changes across different timescales, the first step in DFA is to break the time series into nonoverlapping bins of different bin size. If we start with a small timescale, we could, for example, break the time series into adjacent bins of four data points. As the bins should be nonoverlapping and the time series contains 1,099 data points, this results in 274 subseries, with the last three data points being lost for the analysis on this scale (note that resampling techniques can be used to mitigate such losses; see, e.g., Kantelhardt et al., 2002). Since simple long-term trends that might be present in the data can bias the estimation of monofractal scaling relations (Caccia, Percival, Cannon, Raymond, & Bassingthwaighte, 1997), each subseries of four data points is individually detrended. Usually, this is done by removing linear trends, but one could also choose quadratic or cubic trend functions if these seem more appropriate for the data. After trends have been removed from each subseries, the root-mean-square (*RMS*) for each subseries is calculated across all four-data-point segments (see Equation A1) and averaged to result in the average fluctuation magnitude for that scale (RMS_s). This process is repeated for increasingly larger bin sizes. Figure A2 (see also Figure A3) illustrates the process for subseries with bin sizes of 55, 110, and 275 data points. Usually, the largest bin size corresponds to maximally one quarter of the overall length of the time series.

(*Appendix continues*)

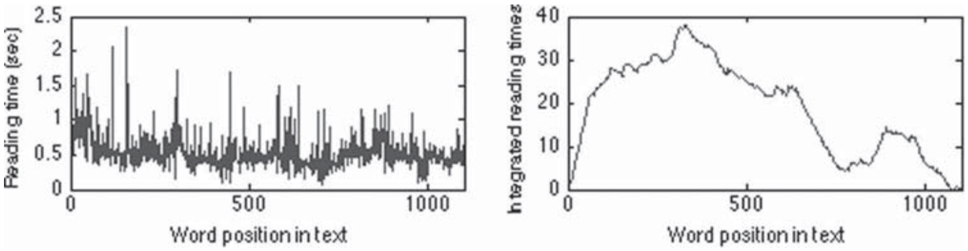

Figure A1. Example time series of one reader's 1,099 word-reading times during text reading. The original time series data are displayed on the left-hand side. On the right-hand side is the integrated version of the time series, which is a preliminary step in the calculation of detrended fluctuation analysis.

$$RMS_s = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[y(k) - y_s(k)\right]^2} \tag{A1}$$

where

RMS_s = root-mean-square for a particular scale.

N = number of subseries at a particular bin size.

$y(k)$ = integrated subseries k.

$y_s(k)$ = local trend for at a particular bin size for the particular subseries k of y.

The slope of that line (S) is the estimate for H, where $H = S$. H is now informative about the strength of monofractal long-range correlations in the data, as depicted in Figure 1. For $H = 0.5$, the variation in the time series corresponds to white noise. For $H = 1.0$, the time series exhibits persistent fluctuations, indicative of strong long-range correlations. In our example, the reading time series exhibits a moderately strong degree of monofractal scaling, as $H = 0.86$ lies between $H = 0.5$ and $H = 1.0$, but somewhat closer to 1.0. For the analysis of the data sets in our study, we chose four as the smallest bin size and one quarter of the maximum length of the time series as the biggest bin size.

Multifractal detrended fluctuation analysis (MFDFA). Multifractal analysis is warranted when the strength of the monofractal scaling relations is not (more or less) homogeneous across the data set, but when H is time dependent. For example, this is often evident in changes between persistent and antipersistent fluctuations in the time series, or bursting behavior (such as can be seen in the time series displayed in Figure 1, around data points 300, 450, 600, and 700). In order to quantify whether there is systematic variation of fractal scaling in the time series, we calculate the multifractal spectrum (*MF*), which gives the range of the different scaling regimes that operate in a single time series.

The calculation of *MF* can be done by a direct estimation, whereby local Hurst exponents are calculated within a single time series and the spread of the distribution of local Hurst exponents estimates *MF*. Here we present an equivalent technique, the so-called indirect estimation of *MF* (Ihlen, 2012), which we used for the data in the present study. The indirect estimation of *MF* proceeds basically in the same way as the monofractal DFA analysis. The difference is that in addition to the standard *RMS* computation, one computes Hurst exponents H for different qth orders of *RMS* (see Equation A2). The standard *RMS* computes the quadratic mean, that is, the average of the squared values. For the q-order *RMS*, the scaling function for the mean of a range of larger and smaller exponents q is calculated as well. Hence, for the standard *RMS*, $q = 2$. In addition to that, we also compute H for exponents ranging from $q = 0.1$ to $q = 3.0$ (see Ihlen & Vereijken, 2010).

$$RMS_{sq} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[y(k) - y_s(k)\right]^q} \tag{A2}$$

where

RMS_{sq} = root-mean-square for a particular scale for a particular q order.

N = number of subseries at a particular bin size.

$y(k)$ = integrated subseries k.

$y_s(k)$ = local trend for a particular bin size for the particular subseries k of y.

q = the exponent of fluctuation magnitude.

The different q-order *RMS* effectively accentuate the magnitude of fluctuations on different scales, where small q accentuates the fluctuations on faster scales (i.e., smaller bin sizes) and larger q accentuates the fluctuations on slower scales (i.e., bigger bin sizes). The rational is that when a time series is composed of homogeneous long-range correlations, then H will not change with variations in q. However, if long-range correlations in a time series are time dependent, then H will be different for smaller and bigger q, and the range of H across different values of q estimates the magnitude of multifractal fluctuations, *MF*.

(*Appendix continues*)

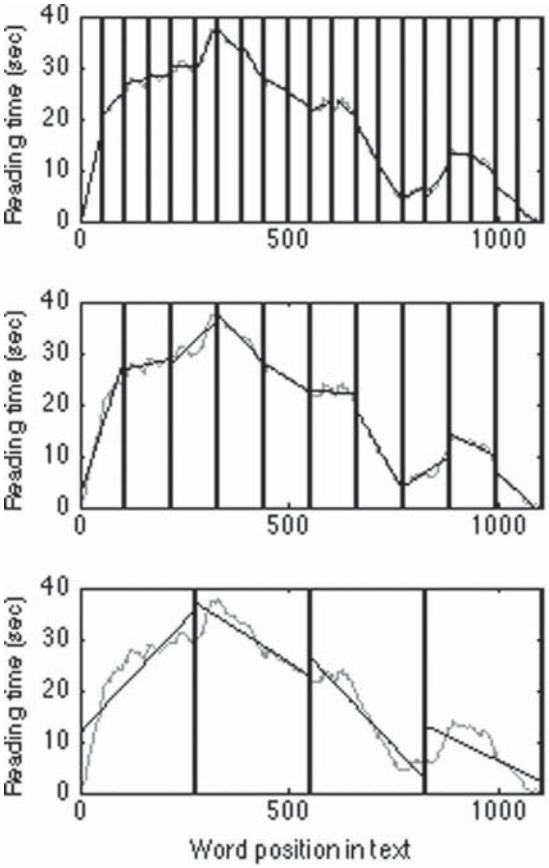

Figure A2. Illustration of the binning and detrending procedure. The time series is broken into increasingly larger bins (consisting of 55, 110, and 275 data points, from top to bottom) to calculate the average magnitude of variation across different scales. After the average root-mean-square (RMS) has been computed over many different bin sizes between 4 and 275, a log-log plot is constructed, charting the logarithm of the average RMS against the logarithm of the corresponding bin size. If the change in average RMS over different bin sizes forms a linear relationship, a scaling exponent *H* (the Hurst exponent) can be computed by fitting a least-square regression line to the plot (see Figure A3).

Figure A4 depicts scaling *H* for the standard value of $q = 2.0$, as well as for the maximum and minimum value of what we employed ($q = 3.0$ and $q = 0.1$, respectively). As can be seen, *H*, estimated by the slope of each regression line on the log-log plot, decreased with increased *q*. From this, the indirect estimation of the multifractal spectrum can be computed (as presented in Figure 2), resulting in a multifractal spectrum width of $hq_{\max} - hq_{\min} = 0.291$ (for details of the computation, and the reason why multifractal spectrum width is not straightforwardly equal to the range of *H* as displayed in Figure A4, left panel, see Ihlen & Vereijken, 2010).

The value of $hq_{\max} - hq_{\min} = 0.291$ does not correspond to the straightforward strength of the multifractal fluctuations. Surrogate tests with shuffled data can be employed to find out more about the actual time-dependent multifractal fluctuations (see Ihlen & Vereijken, 2010; Kantelhardt et al., 2002). Also, the selection of *q* values will influence the actual estimation for the width of the spectrum. Following the recommendations of Ihlen and Vereijken (2010) for relatively short time series of keypress data, we chose *q* values from 0.1 to 3.0 in steps of 0.1. Furthermore, we used the same bin sizes as for the standard DFA, ranging from four to a maximum of one quarter of the time series.

For introductions and best practices guidelines, the following sources can be recommended: For both DFA and MFDFA (as well as related analysis techniques), a recent research topic in *Frontiers in Fractal Physiology* discusses various aspects of applying fractal analysis techniques (Holden, Riley, Gao, & Torre, 2014). Here especially the tutorial introduction on multifractal analysis by Ihlen (2012) can be recommended. Another good introduction to multifractal analysis for psychologists can be found in Kelty-Stephen et al. (2013). A good introduction to monofractal analysis is available on the website of the National Science Foundation (Riley & Van Orden, 2005), as well as in Wallot et al. (2012), who discuss applications to psycholinguistic data.

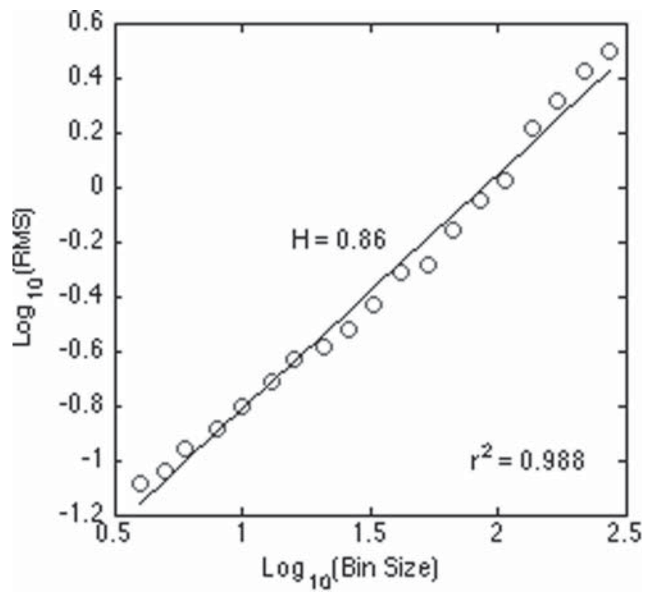

Figure A3. Log-log plot of how the average magnitude of variation (average root-mean-square [RMS]) changes with the scale of variation (bin size). The slope of the regression line fitted to the plot estimates the strength for long-range correlations in the data, captured by the Hurst exponent *H*. The r^2 might appear very high. However, these kinds of log-log plots usually exhibit monotonically increasing functions; hence, an r^2 of lower than .8 or .9 warrants a more detailed investigation of the scaling relation. For example, it might be that there is systematic deviation from linearity on the log-log plot and the data exhibit no clear scaling, or only a subrange of scales actually shows a scaling relation.

(*Appendix continues*)

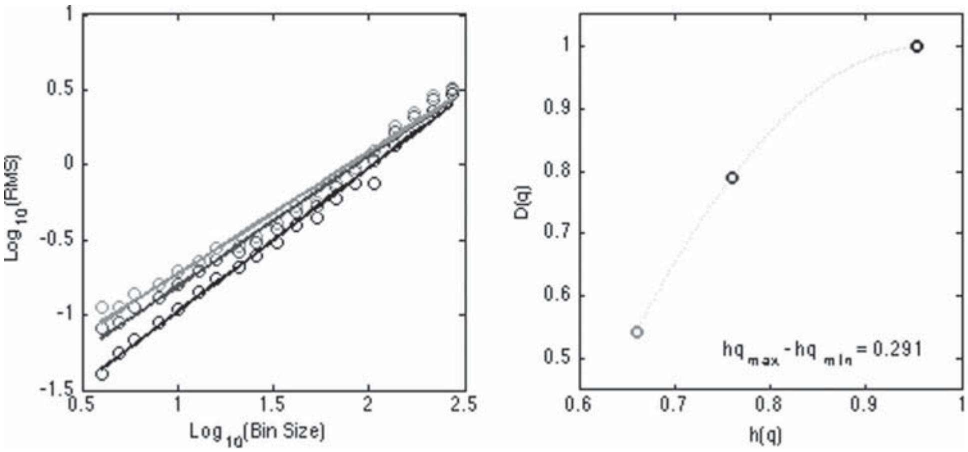

Figure A4. Illustration of the indirect estimation of multifractal spectrum (MF). The left panel displays the scaling functions for three q-order root-mean-square (RMS): $q = 0.1$ (black), $q = 2.0$ (dark gray), and $q = 3.0$ (light gray). As can be seen, the slopes of the lines differ with different q. The right panel displays the multifractal spectrum: The three circles mark the scaling exponents for $q = 0.1$ (black), $q = 2.0$ (dark gray), and $q = 3.0$ (light gray). The dotted line marks the scaling exponents for other values of q that lie between 0.1 and 2.0, and 2.0 and 3.0, respectively. The range of the scaling exponents, $hq_{\max} - hq_{\min} = 0.291$, provides the estimate of MF.

%Determinism of recurrence quantification analysis (RQA). RQA is a rather general class of time series analyses that captures various aspects of stability and instability in time series data (Zbilut, Giuliani, & Webber, 1998). In our case, we were especially interested in the %Determinism measure that quantifies the degree of temporal structuredness and predictability of a time series. To calculate %Determinism for our example time series in Figure A1, a recurrence plot (RP) of that time series needs to be constructed, from which the %Determinism measure can be calculated. As already mentioned in the introduction, RQA uses the method of time-delayed embedding (see Figure 3), whereby a one-dimensional time series is plotted against itself to construct a phase space of that time series. To embed a time series properly, several parameters have to be estimated.

Usually, the embedding procedure begins with estimating a delay parameter (τ), but pertains mainly to continuously sampled signals (Webber & Zbilut, 1994). Keypress series, however, are series of interevent times, and no delaying of the time series is necessary in this case (i.e., the delay parameter τ equals 1). As a next step, we need to find out how often we need to embed the time series, that is, how many dimensions the phase needs to have to properly hold the data (D). This can be done by the false nearest neighbor (FNN) function (Kennel, Brown, & Abarbanel, 1992). The basic idea of this function is that if the phase space of a time series is too small, then data points in that phase space lie close together (i.e., are neighbors) just because the time series has not been embedded properly. This can be solved by seeking an embedding dimension D where data points are sufficiently far away from each other and the number of neighbors stays relatively constant.

To employ the FNN function, the one-dimensional time series is consecutively embedded into higher and higher dimensions and the percentage of neighboring points in each dimension is recorded. Usually, the first local minimum where the number of neighboring points in the phase space stays low across several dimensions is selected as a good estimate for the embedding dimensions parameter. Figure A5 displays the FNN function for our example time series and the first 10 embedding dimensions. As can be seen, the number of nearest neighbors initially deceases with embedding dimension and stays close to a local minimum for D between 3 and 6. As dimension 4 marks the absolute local minimum, selecting 4 for the embedding parameter seems an appropriate choice.

Now the time series can be embedded into a phase space with a τ of 1 and D of 4. Practically, this means that the one-dimensional series x is plotted against itself 4 times, each time with a lag of 1 to construct the four-dimensional vector v that constitutes the phase space (i.e., for our 1,099 data point series, this means that points 1–1,096 are plotted against 2–1,097, against 3–1,098, and against 4–1,099; see Equation A3).

$$v_i = \left(x_i, x_{i+\tau}, x_{i+2\tau}, x_{i+2\tau}, \ldots, x_{i+(D-1)}\right) \tag{A3}$$

where

v_i = is the D-dimensional vector that constitutes the phase space.

X_i = is the ith subseries of the time series x that is used for embedding.

(*Appendix continues*)

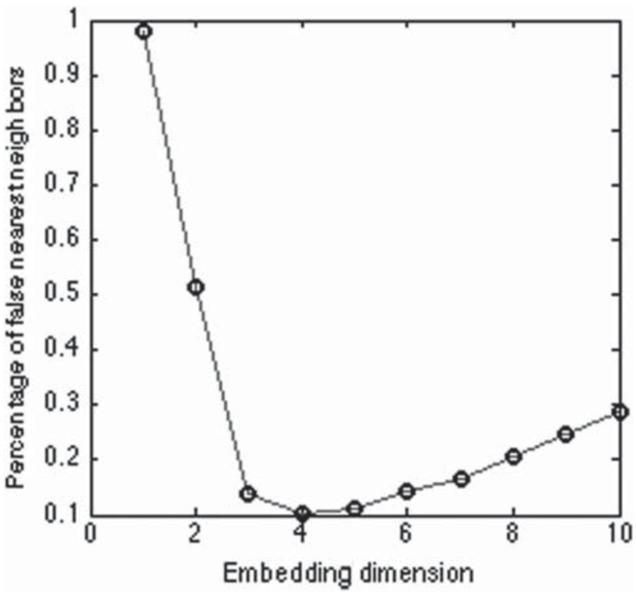

Figure A5. Plot of the false nearest neighbor function for the first 10 embedding dimensions. The first local minimum of the function spans Dimensions 3–6.

τ = is the delay parameter that gives the lag at which the time series x is embedded.

D = is the dimensionality of the reconstructed phase space.

Finally, before we can start to compute RQA, the phase space needs to be normalized by selecting a norm parameter. Normalization of the phase spaces ensures that differences observed between data sets are actually a function of the sequential ordering in a time series, and not due to simple differences in the overall magnitude of the values in that time series. Hence, the most important point about the norm parameter is to keep it constant across all data sets. Because it is less sensitive to single values in the time series, the Euclidean normalization can be recommended as standard choice.

This normalized phase space allows us to create a RP of the time series, which is basically a representation of the distances between points in the phase space. However, the RP is a binary representation: One selects a threshold parameter, and all distances in the phase space that are smaller than the threshold are counted as recurrence, where patterns in the time series are systematically reoccurring, while all distances greater than the threshold parameter are counted as nonrecurrent behavior in the time series. There is usually no absolute value for the threshold parameter for physiological and psychological data. Rather, it is recommended that the threshold is set in a way that 1%–5% of the data points in a time series are counted as recurrent. This way, RQA is sensitive to the systematic structure in a time series (Webber & Zbilut, 2005).

Figure A6 presents the RP of our example reading time series. As a visualization tool, RPs make patterns in time series visible and show structure as a time series repeats itself. RPs are similar to auto-cross-correlation plots, in that they visualize time-lagged dependencies in a time series. Time at lag0 runs along the main diagonal of the plot, stating the simple fact that a time series is always the same with itself at lag0 (i.e., maximally recurrent). Furthermore, the plot is symmetric about the diagonal. The two time series below and to the left of the RP can be used to navigate the plot. For example, when one looks at the coordinate 500 on the *x*-axis and 500 on the *y*-axis, one lands on the main diagonal. When one looks at the coordinate 350 on the *x*-axis and 1,050 on the *y*-axis, one can see that reading time patterns that occurred around word 350 repeat themselves around word 1,050, indicated by the dark area of the plot that is populated with many recurrent points (in contrast, for example, to 25 on the *x*-axis and 1,050 on the *y*-axis, where the white area indicates the absence of recurrent behavior).

A first measure of structure within a time series is %Recurrence, which is related to linear autocorrelation and is simply computed as the sum of all recurrent (black) dots on the plot divided by the size of the plot (see Equation A4). However, one can also see the recurrent points on the plot appear in a more clustered structure, and one measure for this patterning is %Determinism, which is computed as the number of all recurrent points that are diagonally adjacent to each other divided by the sum of all recurrent points on the plot. Hence, %Determinism is a measure of how much reading times are organized in larger temporal patterns that span over sequences of multiple reading times (see Equation A5).

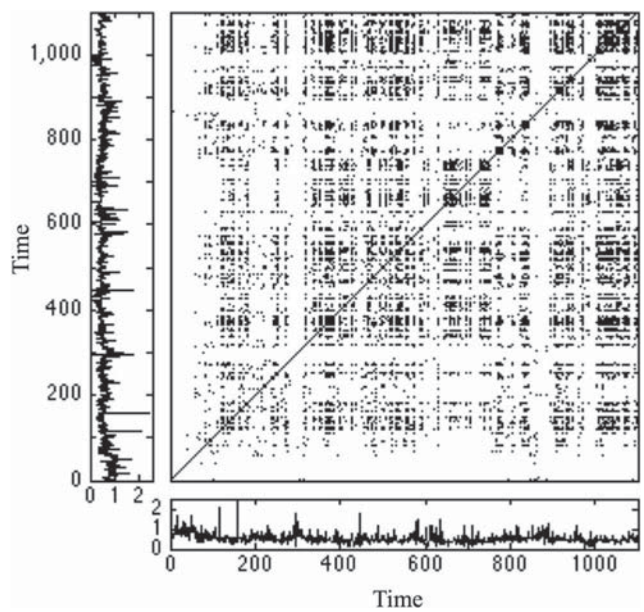

Figure A6. Recurrence plot of the reading time series. The actual time series from which the plot was computed are below and to the left of the plot, making it easy to identify where in the time series reading time patterns recur.

(*Appendix continues*)

$$\%\text{Recurrence} = 100 * \frac{\#\text{ of recurrent points in RP}}{\text{size of RP}} \quad \text{(A4)}$$

$$\%\text{Determinism} = 100 * \frac{\#\text{ of diagonally adjacent points in RP}}{\#\text{ of recurrent points in RP}} \quad \text{(A5)}$$

Our example data set exhibits 7.55% Recurrence and 87.20% Determinism. Similar to the multifractal spectrum, the absolute values of these measures are usually not meaningful in psychological data. Rather, they need be contrasted with an interesting base line or experimental condition—or provide predictive power, such as for comprehension measures of reading. Especially for RQA, the proper parameter setting provides extra challenges for group analysis, as individual data sets will usually differ in optimal delay or dimension parameters. Also, the threshold parameter needs to be set in a way to fit all the data sets in a sample. The standard procedure is to find a single value of delay, dimension, and threshold that fits all data sets reasonably well, and then examine how recurrence variables (e.g., %Determinism) differ across data sets when contrasted on the same parameter settings. Wallot et al. (2012) provide practical advice on how to use RQA with multiple data sets.

For the current study, all parameters were first estimated for each participant's reading time data, and the average values across all participants were used for the analysis, resulting in a Delay = 1 and Dimension = 5. The phase space was normalized with the Euclidean norm.

For further readings on RQA, again the web book on the website of the National Science Foundation can be recommended (Riley & Van Orden, 2005). Also, a repository of introductory guidelines, free software packages, and a comprehensive bibliography of recurrence-plot-related scientific publications can be found in Marwan (2014).

Received December 23, 2013
Revision received April 13, 2014
Accepted April 29, 2014 ■