Nonlinear analyses of self-paced reading

Sebastian Wallot and Guy Van Orden
CAP Center for Cognition, Action and Perception, University of Cincinnati

Nonlinear methods of fractal analysis and recurrence quantification analysis are becoming more commonplace in the cognitive and behavioral sciences. These methods are illustrated here in a tutorial style using self-paced reading data. Self-paced reading was performed in which each spacebar press revealed a story word-by-word or else sentence-by-sentence. Participant readers were either Ph.D. candidates in English literature or undergraduates from an introductory psychology course and the same story was read by all, either one time only or re-read another time on another occasion. The nonlinear analyses revealed crucial differences between the word unit and sentence unit conditions. Performance in the word unit condition was dominated by a task specific strategy, yielding data patterns more like those observed in tapping tasks. Nonlinear analyses of the sentence unit condition, however, discriminated between graduate and undergraduate readers, and first readings of the story from re-reading. From these analyses, the repeated reading of the same story reveals a kind of über-fluency, in a manner of speaking, of the Ph.D. candidates in English literature, whose performance stayed at or closer to a performance ceiling in both readings.

Keywords: fractal analysis, recurrence quantification analysis, time series, reading, complexity, nonlinear methods

Nonlinear analysis of self-paced reading

This article illustrates nonlinear methods of data analyses for self-paced reading data, collected like those reported in Wallot and Van Orden (2011). Three factors were examined in a 2 x 2 x 2 design: the size of the text units presented each time a space bar is pressed — *word units* versus *sentence units*; the fluency of readers — *Ph.D. candidates* in English literature versus *undergraduates* from an introduction to psychology course; and text repetition — *first reading* versus a *subsequent reading*. The nonlinear methods illustrated are timeseries analyses, which give information about the hypothetical system dynamics that could have produced the time-ordered variability observed across reading times. The payoff from these

The Mental Lexicon 6:2 (2011), **245–274**. DOI 10.1075/ml.6.2.02wal
ISSN 1871–1340 / E-ISSN 1871–1375

analyses is new information about performance over and above and different from the results of conventional analyses.

Conventional analyses, in this case, reveal little about the system under study — a hypothetical system that could produce the reading time data. The nonlinear analyses, on the other hand, reveal differences between the word unit and sentence unit conditions in the hypothetical strategy that readers apply during self-paced reading, and discriminate in greater detail between Ph.D. candidate and undergraduate readers in the sentence unit condition. The result is a more nuanced picture of what happened in the self-paced reading experiment, and more nuanced pictures are more usefully constraining of theoretical hypotheses. The promise, then, is a greater capacity to discriminate hypotheses.

We introduce two general categories of methods to estimate nonlinear structure — fractal analysis and recurrence quantification analysis — but before that we describe the methods of the self-paced reading task and the picture of the data that is created by conventional statistical analyses. We then introduce spectral analysis to illustrate fractal methods. We will describe the step-by-step applications of the fractal analyses and then do the same with recurrence quantification methods, ending with a discussion of the results obtained and how they might speak to reading performance. Throughout, we provide references that include more in-depth discussions of the techniques.

Self-paced reading

In the self-paced reading task readers advance through a text at their own pace, which they control using the spacebar of a computer keyboard. Each time a participant reader presses the spacebar the next unit of text appears on the computer screen. The time that passes between key presses estimates the reading time of the text unit. The reading time data of the illustration were sampled from the same story presented either in word units or sentence units. Thus, text accumulated on the screen either word by word or sentence by sentence.

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

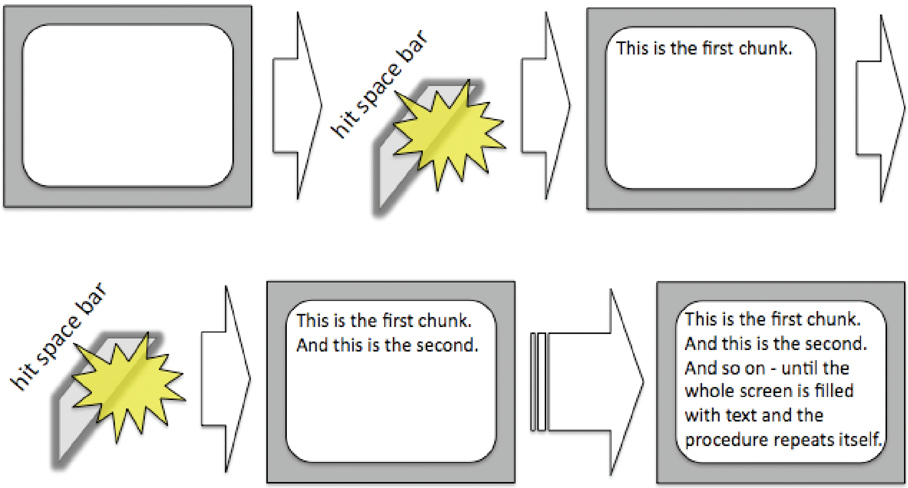

Figure 1. Illustration of the self-paced reading task for sentence-units.

The Arelis Complex consists of 13,930 words, which in turn make up 1,042 sentences. Text units were presented in Times New Roman font (13.5 pt.) on a standard computer monitor, paced by the reader's spacebar presses. Participants read the story while seated before the computer monitor, pressing the spacebar to bring text to the screen. Figure 1 illustrates the self-paced reading procedure.

The key factor examined was the reading fluency of participants, manipulated in a quasi-experimental design. The participants were 32, Ph.D. candidate, graduate students of English Literature and 32 freshman undergraduates attending the University of Cincinnati. We assumed that the graduate students were the more fluent readers, and the data analyses tested for reliable differences in reading performance between graduate student readers and undergraduates. Also, 16 of the graduate students and 16 of the undergraduates read the story more than once.

All participants had normal or corrected to normal vision and reported being unfamiliar with the story. Reading the story took between 30 and 80 minutes, depending on the individual. Participants were compensated for their participation in class credit and cash. After reading the story each reader completed an exam, of which they had been forewarned, to assess story comprehension and memory. The exam required a written summary of the story plot, indicating characters and their roles in the story, plus a multiple-choice sentence completion task. Items for the sentence completion task covered the entire story.

Average reading times

The commonly used methods of data analysis in psycholinguistics estimate parameters such as mean and variance, thought to correspond to population parameters. Contrasts focus on differences among means (and differences among differences to gauge interaction effects). These are all static contrasts however, existing outside of time, even when measured at two different points in time. The nonlinear structures in measured values of behavior are usually manifest over time; they change as they unfold over time. Hence, the analysis of interest must respect the ordered progression of time during the collection of the data. It would matter not all that we randomize the data before computing mean and variance, whereas randomized data lose all information stemming from the order, or time, in which they were collected.

Figure 2 presents average reading times bracketed by standard error bars in the word and sentence presentation conditions. An analysis of variance included the factors *words* versus *sentences*, *graduates* versus *undergraduates*, and *first reading* versus *re-reading*. We observed reliable interaction effects between text unit and reader group, $F(1,56)=6.58$, $p<.05$, and between text unit and repeated reading, $F(1,56)=15.27$, $p<.001$, plus main effects of text unit, $F(1,56)=300.47$, $p<.001$, reader group, $F(1,56)=7.54$, $p<.01$, and repeated reading, $F(1,56)=16.48$, $p<.001$. Words units are read faster than sentence units, graduate students read faster on average than undergraduate students, and a repeated reading of the text yielded faster average reading times than the first reading.

Separate analyses of word unit and sentence unit conditions shed light on the previous interaction effects: When the story appeared word by word, graduate and undergraduate readers produced average reading times that did not differ statistically. There were no reliable differences in average reading times between undergraduate and graduate readers, nor between repeated readings of the text presented word by word, $F<2.08$. However, when the text was presented sentence by sentence, undergraduate readers were slower on average compared to graduate student readers, $F(1,28)=15.91$, $p<.001$. Also, the first reading of the story, sentence by sentence, produced longer average reading times compared to re-reading the story sentence by sentence, $F(1,28)=7.07$, $p<.05$. The apparent interaction effect between reader fluency and number of readings was not statistically reliable however, $F=2.71$.

An analysis of the average standard deviations of reading times also included the factors words vs. sentences, graduates vs. undergraduates, and first reading vs. repeated reading. We observed a statistically reliable interaction effect of text unit with repeated reading, $F(1,56)=12.42$, $p<.001$. The dispersion of reading times narrowed with repeated readings for both reader groups in the sentence-unit condition, $F(1,28)=16.20$, $p<.001$, although the effect was much less pronounced, in fact narrowed to a marginal trend, for word-unit reading times, $F(1,28)=3.17$,

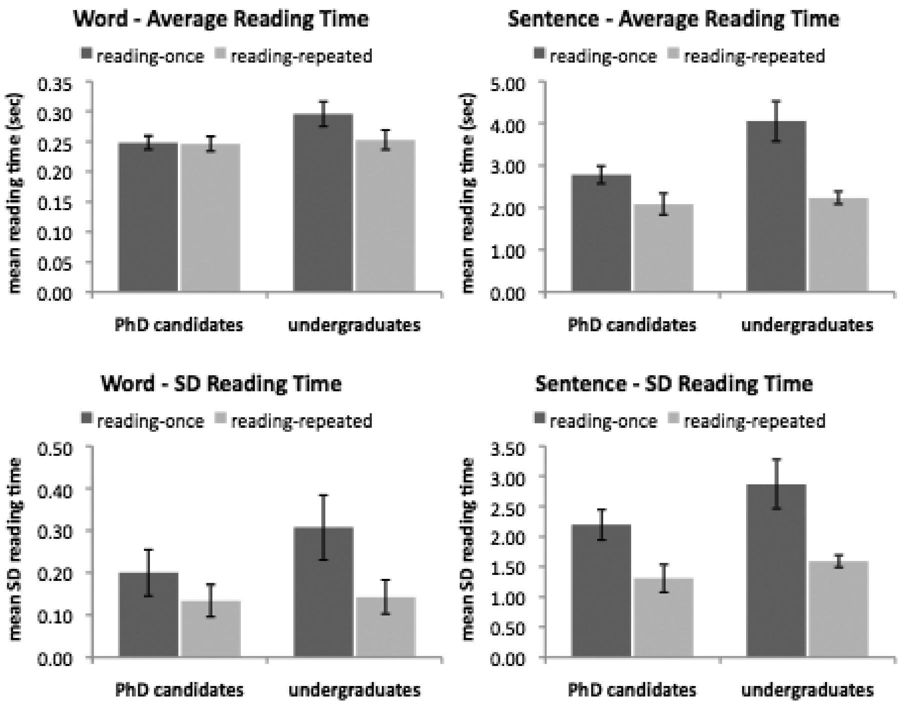

Figure 2. Top panels: Average word and sentence reading times in seconds with standard error brackets. Bottom panels: Average standard deviation of word and sentence reading times in seconds with standard error brackets.

$p = .086$, no other interaction effects were apparent, all $F < 2.35$. Other reliable main effects were observed for text unit, $F(1, 56) = 171.18$, $p < .001$, repeated reading, $F(1, 56) = 19.06$, $p < .001$, and a marginal main effect of reader group, $F(1, 56) = 3.83$, $p = .055$. The primary patterns for interpretation are that dispersion of reading times narrows with repeated readings of the text, and that there is a tendency of graduate readers' reading performance to be less dispersed compared to that of undergraduate readers.

The picture that the conventional analyses present is relatively straightforward. Average performance overall to word units appears to be at ceiling, being close to what might be expected in a simple reaction time task. In that light, the average reading times to word units are relatively unaffected by fluency or repeated readings, suggesting that all the words presented are represented to the same extent in the mental lexicons of all the readers. Thus differences in reading fluency are not due to whether words are represented in the mental lexicon. Perhaps, though, word reading is automatic for almost all words and most of the readers, so performance stays relatively close to ceiling in the word-unit condition. The sentence unit condition opens up the effect of repeated reading, however, and both graduate and undergraduate readers show a reliable decreased in dispersion of reading times as well as faster reading times overall in re-reading the story.

Apparently, the differences based on reader fluency only emerge when the text is presented sentence by sentence. Consequently we might conclude that the locus of reading fluency is in higher order sentence-level knowledge of linguistic structure. Perhaps the knowledge of how lexical items are used in sentences discriminates between graduate and undergraduate readers, and episodic memory of reading a story is sufficient to either instate or compensate for this higher order knowledge. The consequence is seen in faster and more narrowly dispersed reading times.

Fractal analysis

Now we explore what else can be learned about reading performance from the nonlinear structure of the data. One difficulty inherent in the previous picture is that it limits us to inferring differences in reading fluency from quantitative differences in performance, faster versus slower or wider versus narrower dispersion. Difficulty stems from underspecification by the data of the theory that we infer from the data. All we can know is that a quantitative difference exists between fluency groups when the reading units are sentences, but the data cannot tell us about any qualitative differences that are entailed. These must be abducted and then tested using additional manipulations to rule out competing hypotheses, effectively playing a form of 20 questions with nature (cf. Newell, 1973).

Nonlinear methods do discriminate among qualitative differences in performance and, apart from the focus on time evolution, the nonlinear methods make very few assumptions about the system under study, which is an often unappreciated advantage of nonlinear methods. For instance, we need not assume that a stable mean exists, nor that the variance of data is homogeneous, and no assumptions are necessary about the component processes of which a system is composed nor how those processes interact. The first goal, realized up front, is to characterize the form of the interaction among a system's components, whatever their nature (e.g., see Holden, Van Orden, & Turvey, 2009). This means that the way in which we imagine cognition to be, before we design an experiment, matters much less, initially, to the analysis of reading performance. Reading time data can reveal the nature of the interactions among components, for example, and thus tell us how we should proceed past that point (compare Carello & Moreno, 2005).

The method we illustrate requires several hundred data points at least, collected continuously, for reliable quantification of nonlinear structure (see Holden, 2005, for recommendations). Fractal analyses provide two kinds of information to a scientist about the qualities of the data. On the one hand, the results of a fractal analysis tell us whether a conventional analysis can be justified. A fractal analysis estimates the fractal dimension of a data set, and fractal dimensions that differ reliably from $FD = 1.5$ (the fractal dimension of random white noise) may imply that no stable variance exists and sometimes that no stable average value exists. Conventional analyses grounded in the General Linear Model assume stable mean and variance and that data values are statistically independent, one from another, predicting data series that vary in a pattern equivalent to random white noise.

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

The fractal dimension of white noise, $FD = 1.5$, captures this idea nicely if we imagine that white noise leaks halfway into the 2nd dimension. The data graph of white noise is not simply a line, random noise is too irregular to be simply a line, but neither is it a 2-D object, it is "halfway" between. Fractal dimensions exist in between the canonical dimensions of Euclidean geometry. Euclidean dimensions are sufficient to capture the highly regular structures of Euclidean geometry, including 0-D points, 1-D lines, 2-D planes, and 3-D spheres or cubes. Fractal objects live between these canonical objects and the data series presented in Figure 3 is such an irregular fractal object, with $FD = 1.23$. The data graph is not a 1-D line, being highly irregular and aperiodic, and thus leaks somewhat from the 1st dimension into the 2nd dimension, in a manner of speaking.

Fractal dimensions reliably different from that of random white noise imply that a conventional analysis may be unjustifiable and could thus be misleading. A fractal analysis in this case is a test of whether the necessary assumptions of conventional analyses are met. If not, then either the variance of the data series is nonstationary or both the mean and the variance are nonstationary andnonstationary descriptive statistics cannot be trusted to be stable estimates of population parameters (Brown & Liebovitch, 2010).

In particular, the fractal dimension of the timeseries of word reading times from Figure 3, $FD = 1.23$, is very close to a fractal pattern called pink noise (or $1/f$ noise). Idealized pink noise has a fractal dimension of $FD = 1.2$, and thus leaks less into the 2nd dimension compared to white noise. The more important aspects of pink noise for our discussion, however, is that pink noise in timeseries data is an indicator of nonstationary variance. Furthermore, pink noise-type patterns in a

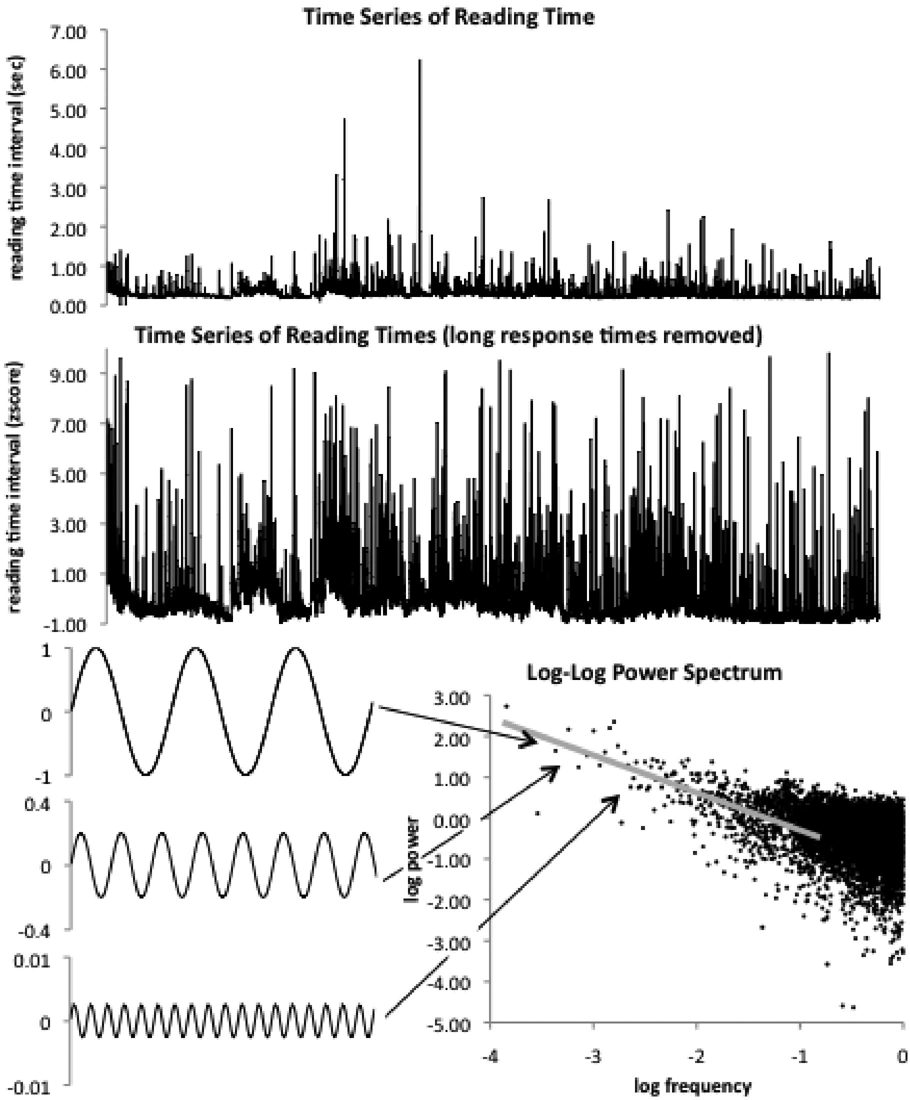

Figure 3. Top: The time series of one participants word reading times. Middle: The same time series, but with extreme outliers removed and normalized is subjected to spectral power density analysis (bottom) where sine and cosine waved are fitted to the data to estimate the power at different frequencies. The Power spectrum is plotted on log-log axis to investigate regions of scaling in the time series.

timeseries reflect that there are tradeoffs between regular and irregular sources of variation that contribute to, or signify the observed performance. It sits between the true randomness and irregularity of white noise ($FD = 1.5$), and the perfect regularity and orderliness of a one-dimensional line in the absence of variation.

When these two aspects are in balance, a pink-noise scaling relation can be observed, which is neither over-random nor over-regular.

In terms of reading performance, we could imagine that a text is not simply a source of regularity and order for the reader, but also a source of disorder. For example in a self-paced reading task that reveals a text word-by-word, each word provides a definite opportunity to respond (in terms of the recorded key-presses) and gives guidance to comprehension of the content as the text unfolds. This might be a source of regularity, this ordering aspect of the text. At the same time, the opportunities to respond to — and comprehend — each word are somewhat contingent on the course of the text up to the point of the upcoming key-press as well as situational or idiosyncratic aspects of the reader. These aspects of self-paced reading may be sources of randomness.

In over-regular performance, participants' key-press intervals would mimic a stable text or word property, in over-random performance, key-press intervals would be completely unaffected by the text presentation or the text would not possess ordering characteristics that key-pressing could entrain to. Fractal statistics of a timeseries capture the tradeoff between these tendencies, yielding a fractal dimension that can be brought under experimental control. For example, one aspect of our own work has been to examine and interpret the changes in fractal dimension due to differences in task demands, participant characteristics, and experimental manipulations (e.g. see Castillo, Van Orden, & Kloos, in press; Kloos & Van Orden, 2010; Van Orden, 2010; Van Orden, Kloos, & Wallot, 2011). As we illustrate shortly, the present self-paced reading data also present different fractal dimensions in different conditions.

In addition to the two kinds of information supplied by fractal analysis, there are presently two kinds of fractal analysis applied productively to human event time data. A monofractal analysis estimates a single fractal dimension across an entire data set. Monofractal analysis assumes that a data series has a stationary fractal dimension. This assumption is not always true however. Human response time data are sometimes monofractal but more often the fractal dimension is nonstationary, changing reliably during the course of data collection.

Multifractal analysis tests whether a data set has a stationary fractal dimension and whether the dynamics of a performance are better described by multiple fractal dimensions, varying across a multifractal spectrum. Multifractal analysis is an extension of monofractal analysis. It assesses specifically whether the variation in fractal dimension across a data series is substantially different from what would be expected of monofractal data. Thus multifractal analysis tests whether different qualities of variation exist in the same data set. A positive outcome implies that the same data series entails multiple different organizations of the system that performs a task.

Early work using fractal methods in behavioral science was often modeled after studies in physiology (e.g., Bassingthwaighte, Liebovitch, & West, 1994). The work in physiology has culminated recently in the journal *Fractal Physiology* of the *Frontiers of Science* series. Holden (2005) is a tutorial introduction to spectral and standardized dispersion analysis to estimate a single fractal dimension of a data set. Peng, Havling, Stanley, and Goldberger (1995) describe another monofractal method called detrended fluctuation analysis applied to heartbeat data. Ihlen and Vereijken (2010) discuss multifractal analyses of human data series, including time estimation data and word naming data. Seuront (2010) is a good text book although its examples come from ecology and aquatic science.

Monofractal analysis

Figure 3 illustrates a spectral analysis of one participant's repeatedly measured word by word reading times. As noted already, we can expect to satisfy two goals with the results of a monofractal analysis. One result can decide the extent to which we were justified in drawing conclusions from the previous conventional contrast of averaged word reading times. The second result concerns whether fractal statistics can reliably distinguish graduate student from undergraduate readers or a first reading of the story from a subsequent reading.

Data preparation

We followed the guidelines of Holden (2005) in preparing the reading times for the spectral analysis. Although outlier reading times are in principle legitimate data values in nonlinear analyses, sufficiently extreme outliers can bias a spectral analysis, possibly prompting a false rejection of conventional analysis. Thus, extreme outliers were removed to ensure reliable conclusions in that regard. Word reading times less than 100 ms or greater than 2,500 ms were eliminated from the spectral analysis, as well as sentence reading times less than 100 ms or greater than 20,000 ms.

Detrending also reduces bias toward a false rejection of conventional analysis, and if the pattern of reading times is truly fractal then the removal of outliers and trends should not grossly impact the results (Caccia, Percival, Cannon, Raymond, & Bassingthwaighte, 1997). A safe bet then is to conduct the analysis with and without outliers and trends and trust similar outcomes. In the second step, all the reading times that fell outside of three standard deviations from a participant's average reading time were removed, as well as linear and quadratic trends.

All the available techniques to estimate fractal dimension are workarounds of linear methods of analysis and are thus prone to artifacts. To minimize artifacts we use several fractal methods together, checking one result against the others. Standardized dispersion analysis is preferred if the scaling exponents of the data lie between white and pink noise (that is, between fractal dimensions of 1.5 and 1.2, respectively, see Holden, 2005). Also the same trimming and detrending should be applied across the analyses to produce comparable results (even though detrended fluctuation analysis does not require removal of trends for example, see Peng et al., 1995).

Holden (2005) recommends trimming from the beginning or end of the data series down to an integer power of two (e.g., 1024, 2048, 4096, …) to make the analysis more efficient computationally and to insure that all data sets are of similar length. We have skipped that recommendation because the computational power of available platforms eliminates most concerns about overly long running times and because the present reading time data series have visibly different dynamics initially (e.g., strong initial transient phases in reading times), which we wished to include in the analysis. Finally, the trimmed and detrended data were subjected to spectral analysis, using the Fast Fourier Transformation.

The spectral analysis of each participant's reading times yielded a log-log plot like that portrayed in Figure 3. The Y-axis in the plot tracks the magnitudes of variation as changes in reading times and the X-axis tracks how often changes of particular magnitudes occur. The relation between size of change and frequency of change is estimated by a least-square regression line, which quantifies the relation between size and frequency in the slope of the regression line. The end result estimates a scaling relation between size (S(f)) and frequency (f): $S(f) = 1/f^{\alpha}$.

The slope estimated in Figure 3 is negative, $-\alpha = -0.76$, which in turn gives us a scaling exponent $\alpha = .76$. This value of the scaling exponent, $\alpha = .76$, lies between the scaling exponents that correspond to idealized white noise (scaling exponent of $\alpha = 0$, slope = 0, fractal dimension = 1.5) and the idealized fractal pattern that we referred to as pink noise (scaling exponent of $\alpha = 1$, slope = −1, fractal dimension = 1.2). The scaling exponents from all conditions can be subjected to conventional inferential statistical analysis (a two-way ANOVA in the present case) to compare average scaling relations from different conditions.

The absolute scaling exponents are of interest to determine whether conventional analyses can be justified. In the example presented in Figure 4, the scaling exponent is within the range of scaling exponents between white noise and pink noise. This range is associated with fractional Gaussian noise, which has no defined variance; variance will increase with increases in sample size (see Van Orden et al., 2011; Wallot & Van Orden, in press, for examples). Fractional Gaussian noise has a stable mean that can be informative depending on the hypotheses that are tested.

Otherwise, like other descriptive statistics, scaling exponents (and fractal dimensions) are most informative in contrasts that compare different outcomes, and laboratory performances can yield a wide range of scaling exponent values (Van Orden et al., 2011). Also, just as for other performance measures, the observed variation is due to tradeoffs among task demands, the integrity of the system, participant skills, experimental manipulations, and participants' strategies (Kloos & Van Orden, 2010; Van Orden, 2010).

Results and discussion of monofractal analysis

The analysis of variance yielded no statistically reliable interaction effects (all F values < 1.7, all p values $> .20$) across the factors *words* versus *sentences*, *graduates* versus *undergraduates*, and *first reading* versus *repeated reading*. The average scaling relations of reading times are portrayed in Figure 4. The most pronounced difference is between word-unit and sentence-unit presentation conditions, $F(1, 56) = 78.98$, $p < .001$. In general, changes in α, in the direction of the $\alpha = 0$ of white noise, have been associated with an increase in sources of involuntary control of task performance (or a reduction in voluntary control, Van Orden et al., 2011, is a review). Keep in mind, though, that the α values close to zero are yet reliably different from zero, indicating the presence of fractal behavior and not simply random variation (Gilden, 2001, 2009; Holden, Choi, Amazeen, & Van Orden, 2011).

The difference between word unit (mean $\alpha = .55$) and sentence unit (mean $\alpha = .19$) presentation conditions is a difference in task demands. Task demands are sources of involuntary control, suggesting that sentence unit presentation conditions increase the influence of task demands on performance measures. If the story presented had had a rhythmic structure, like some poetry, then the change from sentence to sentence could have presented systematic changes that we might have picked up as systematic changes in reading times. Otherwise, effectively, unsystematic changes in sentence structure and content are unsystematic perturbations to the measured values that are input to the spectral analyses, and unsystematic perturbations change the estimates of α values in the direction of random white noise.

Task demands usually vary unsystematically from one measurement trial to the next unless the experimenter designs things otherwise. Measured values that resemble somewhat white noise or values that change in the direction of white noise reflect the increased presence of unsystematic changes in task demands, as the ordered series of measured reading times varies more like a series of randomly selected reading times. Unsystematic differences among sentences are the likely source of the more random pattern, assuming that the sentences that compose the story vary in length, complexity, and other linguistic properties, and vary unsystematically with respect to their reading times in the order in which the sentences appear in the story.

But why would sentences vary more randomly in this regard than words? The syntactic, semantic, and surface properties of the words that compose the story's sentences also vary unsystematically with respect to word reading times. This is not to say that word order does not matter in sentences, only that word order is unlikely to be more systematically related to reading times than is sentence order. In fact fractal analyses of story-ordered word properties yielded α values close to white noise, namely $\alpha = -0.03$, suggesting that word-order contributes unsystematic variation. These results render an explanation in terms of the difference in task demands alone insufficient.

If the difference between word and sentence conditions is not exclusively in the linguistic or psychological properties of text units, then it may be in the strategy that the reader brings to the task — it may be that the word unit condition also encourages different voluntary control of task performance (rather than a change in involuntary control exclusively). For instance, words are read and understood almost immediately, and more quickly than the time it takes to initiate and follow through with pressing the space bar. Skilled silent reading by undergraduates reaches average speeds of around 300 words per minute (cf., Rayner & Pollatsek, 1994) — close to, simple reaction times, and faster than typical performance in a tapping task in which undergraduates produce repeated key presses at their self-selected comfort pace (McAuley, Jones, Holub, Johnston, & Miller, 2006).

An efficient strategy in the word unit condition might be to find a fast pace for repeatedly pressing the space bar and simply read the words apace as they appear across the screen. This strategy explains the data nicely because the average α value in the word unit condition is within the range of values found in tapping performance (Chen, Ding, & Kelso, 1997; Deligniéres, Torre, & Lemoine, 2009). The hypothesis is also consistent with the statistically reliable changes due to number of readings of the story, $F(1, 56) = 5.57$, $p < .05$, together with the absence of reliable differences between graduate and undergraduate readers, $F < 1$.

The latter result makes sense because graduate and undergraduate readers would not obviously differ in a tapping strategy, and the former result makes sense if the strategy is more reliably engaged in repeated readings when both graduate and undergraduate readers are familiar with the story. For instance it is possible that less familiar words or complicated passages would increase word reading times intermittently but sufficiently in the first reading to supply unsystematic perturbations of measured values. A subsequent re-reading of the story could dampen these perturbations as repetition of story elements yields consequently faster comprehension of the now familiar words and sentences.

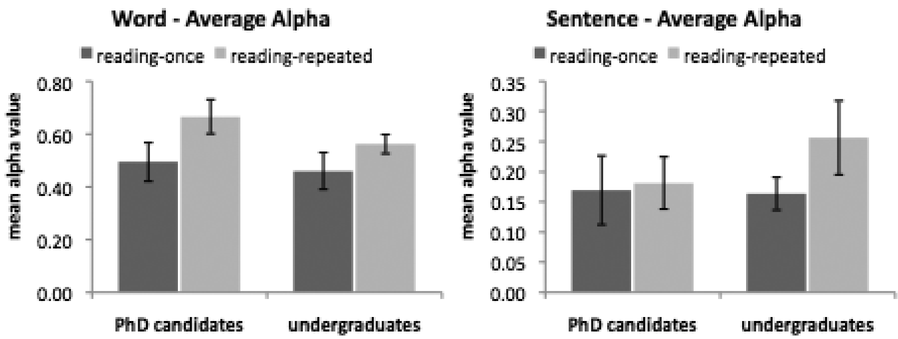

Figure 4. Average α values for word and sentence reading times bracketed by their standard errors.

The previous ideas gain credibility in a closer look at reading performance, in the separate analyses of *word by word* versus *sentence by sentence* presentations of the story. The estimated α values across word-by-word reading times are portrayed in the leftmost bar graph of Figure 4. An analysis confined to the word unit condition reveals a reliable change in average α values with repeated readings of the story, $F(1,28)=4.79$, $p<.05$. The average α values change from values closer to white noise ($\alpha=0$) toward values closer to pink noise ($\alpha=1$), and the same pattern of change occurs reliably for both graduate and undergraduate readers, which do not differ statistically otherwise, and there is no statistically reliable interaction between these factors, both $Fs<1.22$. These changes are unique to the word unit condition because there are no reliable differences among the much smaller α values of sentence reading times, all $F<1.14$. It makes sense then that word unit presentation encompasses a tapping strategy in self-paced reading.

Multifractal analysis

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

Mathematically, the tools of multifractal analyses are formal extensions of the tools of monofractal analysis. The monofractal α value can even be vaguely thought of as the central tendency of the multifractal outcome, and the spectrum of α values that result from a multifractal analysis can be seen as a dispersion of α values around that central tendency. A key statistic of interest is thus the width of the multifractal spectrum (as in a loose analogy with a standard deviation or standard error of a mean). If the width is sufficiently small, the monofractal α value is a good enough estimate of the fractal properties of a data set. If the multifractal spectrum is sufficiently wide then it would be a mistake to ignore this essential character of the data, that its fractal properties are changing during the task performance. Fractal properties that change indicate that the underlying dynamics are changing in quality (Riley & Turvey, 2002).

Multifractal phenomena are unabashedly complex in the strict formal sense of complexity science, implying interaction-dominant dynamics among the component processes of the reading performance (Ihlen & Vereijken, 2010). The presence of multifractal dynamics in data may also imply that a studied system self-organizes its behavior to suite the requirements of the task. The implications of this chameleon-like capacity come from the fact that the only known systems that produce multifractal data are systems that self-organize their behavior in interaction-dominant dynamics. Interaction-dominant dynamics allow the component processes of a system to change each others dynamics perpetually as they interact (Van Orden, Holden, & Turvey, 2003). This creates interdependence among component processes, which allows context to constrain changes in all interacting processes, becoming constitutive of the self-organization (Van Orden, Kello, & Holden, 2010).

Results and discussion of multifractal analysis

We conducted multifractal analysis of the reading time data, applying two prominent analysis techniques. Multifractal Detrended Fluctuation Analysis (Kantelhardt, et al., 2002) and Multifractal Continuous Wavelet Transformation (Ihlen & Vereijken, 2010) were used together, allowing us to compare outcomes and enhance statistical reliability by ensuring common outcomes. We report the average widths of multifractal spectra estimated using the multifractal continuous wavelet transformation. The single statistically reliable interaction was between *word vs. sentence* unit presentations and *single vs. repeated reading* of the story, $F(1,56) = 9.69$, $p < .05$; all other interaction terms yielded Fs < 1.29, $p > .25$. Figure 5 portrays the average widths of the multifractal spectra in the different conditions. The average width of the multifractal spectrum from word-unit reading times shrinks reliably with repeated readings for both reader groups, $F(1,28) = 8.92$, $p < .01$. There was no effect of repeated readings or reader group on the multifractal spectrum of sentence unit reading times, all $F < 1.53$.

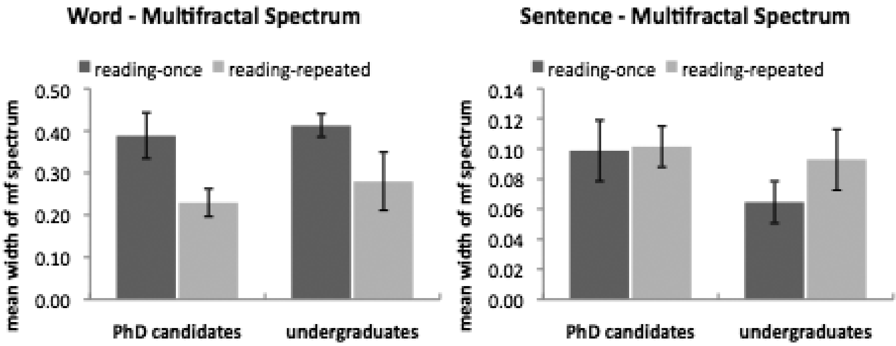

Figure 5. Average widths of multifractal spectra computed from stimulus-ordered series of word and sentence reading times, bracketed by the standard errors of the means.

Fractal and multifractal characteristics of sentence reading times are not reliably different for different reader groups or repeated readings. Monofractal exponents are close to white noise in the sentence-unit condition and the widths of the multifractal spectra are narrow. Thus sentence reading times are fairly well characterized by their monofractal exponent, which suggests the presence of unsystematic perturbations to reading times due to unsystematic variation of sentence properties, in the story order of sentences.

However, both the monofractal and multifractal analyses are sensitive to changes in the pattern of word reading times. Monofractal exponents increased with repeated readings and the width of the multifractal spectrum decreases with repeated readings. Both these outcomes make sense with respect to the hypothesis that the word unit condition allows a special tapping strategy, to tap quickly and read words apace. Again, graduate and undergraduate readers would not obviously differ in a tapping strategy, and the strategy is more reliably engaged in repeated readings when both graduate and undergraduate readers are familiar with the story. When the story is familiar, the influence of story idiosyncrasies is reduced allowing a more coherent tapping strategy and a fractal pattern closer to pink noise, deviating to a lesser degree from pink noise reflecting a tighter balance between voluntary and involuntary sources of control.

Recurrence quantification analysis

Recurrence quantification analysis begins with a phase-space reconstruction. The phase space portrait is inferred from the 1-D data series of reading times. This phase space is usually higher in dimension than the 1-D data series. A vector of the ordered reading times is thus used to plot reading times against themselves, using one (or more) constant lag(s) or delay(s) yielding the higher dimensional portrait of the system dynamics. The higher-dimension phase space is used, in turn, to define structure in a 2-D recurrence matrix, in which only equivalent reading times at equivalent locations in the phase space are marked. Webber and Zbilut (2005) is an accessible introduction to phase-space reconstruction in RQA; the mathematical rationale and proofs can be found in Takens (1981).

The recurrence matrix is formed by following the order in which the data points were collected on both axes of a matrix, and then marking each point in the matrix at which equal reading times recur at the same location, along a shared trajectory, in the higher-dimension phase space. The recurrent reading times are thus located in the neighborhood of equivalent values along a shared trajectory of the space that was formed using lagged reading times.

Recurrence quantification analysis (hereafter RQA) quantifies the geometry of the recurrence matrix. In doing so, RQA yields variables that quantify the dynamics of a hypothetical system that could have produced the reading time data. The central concept of recurrence concerns how often a system repeats its behavior. Data that follow a regular pattern of variation repeat the same behavior often, producing a lot of recurrences and a simple regular geometry in the recurrence matrix. In more complex data series, recurrences become less frequent and the pattern of recurrences itself can change across the recurrence matrix. Recurrence quantification analysis quantifies these aspects of the matrix geometry producing variables that can distinguish among different data sets.

The number of dimensions, estimated from reading times, also estimates the number of active system components (i.e., active degrees of freedom) producing the variation in reading times. RQA reconstructs the latent higher-dimensional variation from 1-D data. It works because continuously or instantly updating feedback creates coupling among the active degrees of freedom sufficient to create interdependence among the components. Interdependence means that each change in each component is reflected simultaneously among all components. Interdependence thus imbues the changes in each component with information about how the other components are changing. Because everything is connected to everything else during behavior we can use the information latent in 1-D data to recover the higher-dimensional dynamics of the hypothetical dynamical system in question.

Marwan (in press) discusses contemporary issues in applying the analysis, including how many data points are necessary to conduct Recurrence Quantification Analysis analysis properly. The range is wide: sometimes, several hundred data points suffice, sometimes tens of thousands of data points are needed, depending on the complexity of the data series. As a general rule, if the results of a recurrence analysis are based on less than 400 data points, they should be interpreted with caution and appropriate skepticism.

Data preparation

In a report of RQA analyses, it is import to describe how the data were prepared before the analysis, how many data points each data series contained, the parameter values that were picked, and which parameters were held constant across the analysis. The data should be normalized to begin with, setting the parameters *norm* and *rescaling*, usually to the Euclidean norm and the maximum distance rescaling option. Apart from special circumstances, the most important thing about norm and rescaling is to keep them constant across the analyzed data sets that are compared.

Also, RQA mandates the presence of each reading time, kept in the same order as the reading times were collected. If any data points are eliminated then their adjacent reading times, on either side, are moved closer together. This distorts the phase-space portrait, which distorts the recurrence plot that, in turn, distorts the quantities computed from the recurrence plot. Hence, only non-trimmed and non-detrended data series are subjected to RQA — with one exception:

Suppose that a recorded data series contains data from two different contiguous events, for example eye-position measures recorded by an eye-tracker during calibration versus eye-position measures taken right after, during a reading experiment. In such a case, the calibration part of the data series can be removed up to the point where the eye-movements of reading begin. If the data series is incomplete (that is, it is truly missing data points), it is advisable to estimate the distribution of the data series and randomly draw values from the distribution to fill in the gaps. It would not be advisable to insert systematic values for missing data points, such as the averaged values of adjacent data points. Using averaged values risks inserting spurious structure into the recurrence plot.

Figure 6 is a cartoon illustration of a 1-D data series converted into a 3-D phase-space portrait. The data plot in the upper-left quadrant of the figure has time (presentation order) on the X-axis and the data measurement on the Y-axis. The data plot is also marked with three dots to indicate the lag delay for re-graphing the data in a 3-D portrait. The delay parameter is the distance, on the X-axis, between the lagged points. The lagged points are used to re-plot the data series (against itself) in the higher-dimension phase space.

The delay parameter is selected so as to minimize the interaction between points of the measured time series (maximize the information available across the lagged points). This parameter is commonly estimated using the first minimum of the autocorrelation function or else using the mutual average information function of the data series. With the delay parameter in the example set at four, the data series can be replotted. The Y-value of the left-most dot in Figure 6 defines the value on a new X-axis in 3-D, the Y-value of the middle point is the value on a new

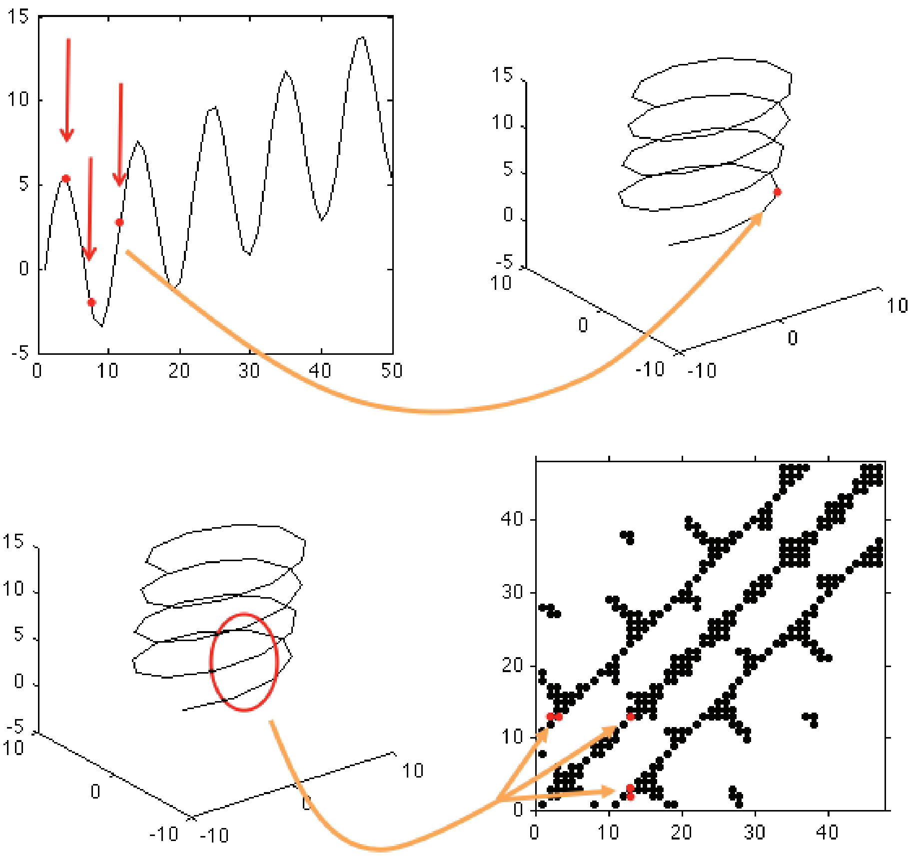

Figure 6. From a one dimensional time series to a three-dimensional reconstructed phase-space to a recurrence plot. The top row illustrates how a time series plotted against itself at time t4, t8, and t12 yields coordinates for a three-dimensional phase-space portrait of that time series (i.e., x = t4, y = t8, z = t12). The bottom row illustrates how a neighborhood (indicated by the red circle) in the phase-space gives rise to one or more points in recurrence plot.

Y-axis, and the Y-value of right-most point is the value on a new Z-axis, resulting in the (x,y,z) point in 3-space pictured in the upper-right quadrant of the figure.

The dimension parameter defines the number of dimensions of the phase-space portrait that is (re-)constructed (the number of times the time series is plotted against itself with the specified delay). The minimum of a nearest-neighbor function is usually used to estimate the number of dimensions needed. The data points that are used in this procedure (delay * times dimension) are lost for the quantifications. Hence if dimension is a high value and or delay is a large value, then the remaining data points have to be sufficient in number to reliably estimate quantities from the recurrence plot.

With all the data replotted in 3-space (the spiral shape of the upper-right quadrant), a neighborhood criterion is chosen — the radius parameter — portrayed in the lower-left quadrant by a circle. The radius parameter sets a threshold within which points in the reconstructed phase-space portrait are counted as equivalent recurrent points. The radius is chosen conservatively to define the equivalence neighborhood around each trajectory in phase-space, which decides which nearby points should be counted as intersecting trajectories (given a little noise). Generally, the radius parameter is chosen so that the recurrence rates do not exceed 5%.

Usually, the radius size is kept constant for all the data sets that are compared. However, it is sometimes more useful to keep recurrence rate constant, letting the radius vary. This is particularly useful if no radius can be found that brings all the data sets into the range of 1–5% recurrence rate. In Figure 6, the three, vertically stacked, data points, located one above the other, within the radius of the middle point, are deemed equivalent (with the middle value) — they are recurrent points on intersecting trajectories. Consequently, when we plot the recurrence matrix a dot can be inserted to mark all possible pairs as recurrent points.

Five recurrent points are illustrated in the recurrence matrix in the bottom-right quadrant of Figure 6. The recurrence plot is fully fleshed out by marking each intersecting point of recurrent values with a dot. With all these dots in place the larger structures of the recurrence plot become visible as lines of dots, 'empty spaces,' and larger geometrical structures. The visible structure provides the basis with which to quantify the nonlinear variables of the recurrence quantification analysis.

For the reading time data, the entire data series of 13,925 data points for words and 1,042 data points for sentences was subjected to RQA without trimming or detrending. The data were transformed into z-scores prior to the analysis. The parameters were estimated for each data set individually. Mutual-average information was used to estimate delay parameters ranging from one to 54, and false-nearest-neighbor analysis estimated embedding dimensions ranging from three to 11.

We used the same delay value = 2 for all data sets, dimension = 8 for word-unit readings and dimension = 4 for sentence-unit readings. The radius parameters were set to yield averages of 2% recurrence for each text unit. For data sets that are characterized by roughly the same parameters, it makes little difference whether each data set enters the analysis with its individual parameter settings, or with parameter values chosen to be identical for all data sets.

RQA yields multiple quantitative outcomes variable, and more are currently under development. Not all outcome variables are always applicable or meaningful for every data set (see Marwan, in press). Hence, one must pick measures that make sense in the context of the data at hand and the research questions to be answered. Some measured values such as recurrence rate (%RECurrence) or the amount of deterministic structure in the data (%DETerminism) are meaningful in most contexts, and for this illustration using the reading data we chose %REC, %DET, and TrappingTime, to illuminate self-paced reading.

RQA of reading times

Figure 7 portrays %RECurrence, %DETerminism, and TrappingTime results for word and sentence reading units. No reliable main effects or interactions were observed for repeated readings or reader group, all $F < 1.33$, for word response times. However, the sentence unit data yielded a marginally reliable interaction effect between repeated readings and reader groups for %DET, $F(1, 108) = 2.93$, $p = .098$, TT, $F(1, 108) = 3.15$, $p = .087$, and the same pattern is replicated in %REC.Recurrence variables each provide independent information about a data series, so similar patterns are replications, adding significance to visible yet marginal effects. The present pattern in %REC, %DET, and TT is a decrease for graduate readers in a repeated reading, while the same factors increase for undergraduate readers in a repeated reading.

Higher %RECurrence (for the same value of the radius parameter) indicates that the behavior is less spread out in the phase space, perhaps less noisy. %DET equals the number of diagonally adjacent points in the recurrence plot divided by the total number of recurrent points, which estimates the degree of order in the data, the extent to which readers fall into coherent trajectories of similar reading times across text units. TT is the average time that the system stays in (i.e., 'is trapped in') one state, an epoch of very slowly and gradually changing reading times, for example.

The contrast between results from word and sentence unit conditions corroborates the hypothesis that word by word reading is of a different quality than sentence by sentence reading. The word unit condition resulted in similar outcomes for undergraduate and Ph.D. candidate readers, consistent with the hypothesis that they find a common strategy of quick taps of the spacebar while reading words apace. The cross-recurrence quantification analysis will provide another, stronger test of this hypothesis.

The sentence-unit condition can now be seen to be the more interesting condition because it distinguishes between Ph.D. candidate readers and undergraduate readers (re-reading a text has apparent opposite effects on the two categories of readers across the RQA variables). The opposite directions of the effects may imply different qualities of effects and different qualities of readers — fluent undergraduate readers versus über-fluent Ph.D. student readers.

There is another way to test this hypothesis using cross recurrence analysis. Cross recurrence analysis can be used to look at commonalities among readers in

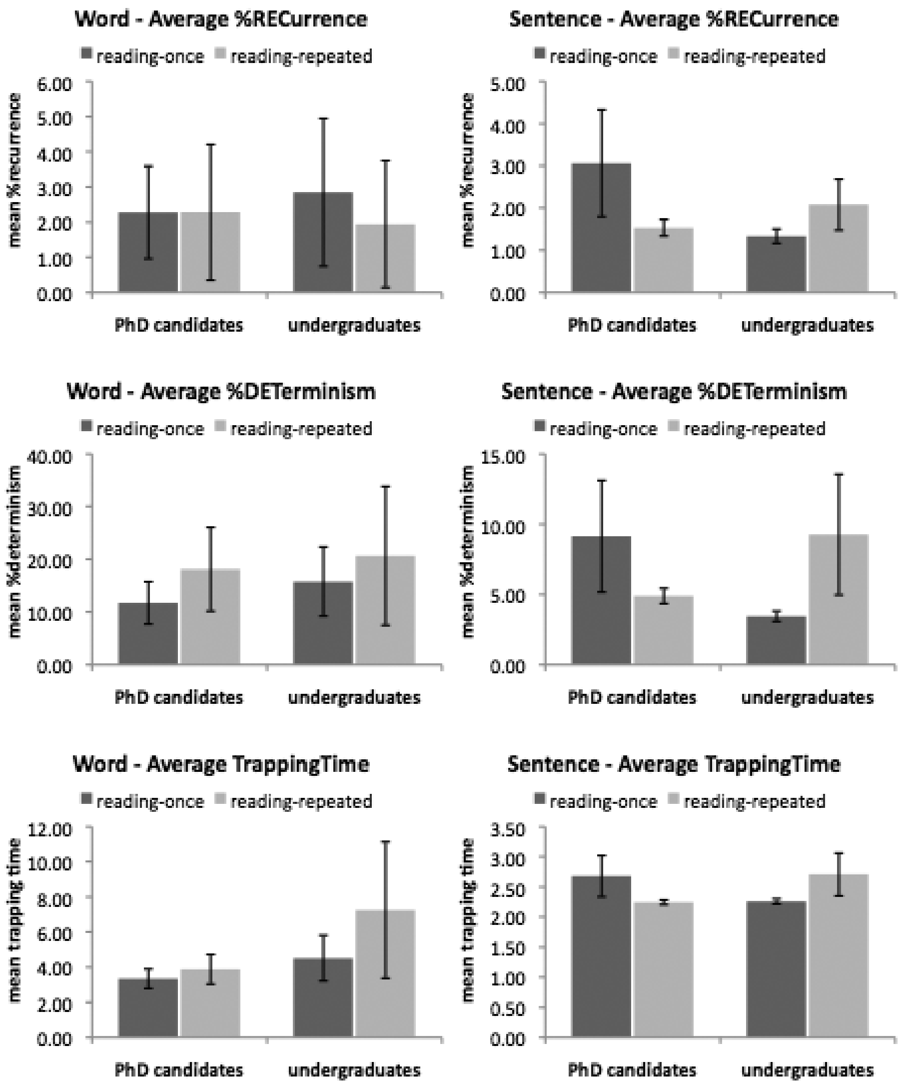

Figure 7. %RECurrence, %DETerminism and Average TrappingTime, with standard error bars, for word and sentence presentations conditions.

a group that are due to the common accommodation of task demands, including how the text-unit conditions affect readers similarly or differently.

Cross-recurrence quantification analysis

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

The procedure for conducting the CRQ analysis is nearly the same as that for conducting RQA, the primary difference being that two data sets are contrasted instead of only one. That being the case, it can be crucial to normalize both data sets prior to analysis (for example by computing z-scores) so that they share a common scale. Shockley (2005) gives a good introduction to cross-recurrence quantification analysis (and see also Marwan, in press). We used CRQ to compare the shared dynamical structure among readers who read the same text under the same conditions. So, for instance, we asked whether two undergraduates produced shared patterns reading times because they read the same text, advancing it identically, unit by unit in spacebar presses. The idea behind this analysis is that readers may possibly entrain to a text in self-paced reading or adopt the same strategy for advancing the text in the self-paced reading task. Such commonalities may yield parallel dynamics that can be captured in CRQ.

Each condition included eight readers and we examined all possible pairings, which yielded $8!/(2! * (8-2)!) = 28$ possible pairings in each cell of identical reading conditions. The data were transformed into z-scores prior to the analysis. The parameter settings for the CRQ were identical to those chosen for the RQA (delay value = 2 and dimension = 8 for all data sets, including both word and sentence data, radius parameters were set to yield averages of 2% recurrence for each text unit).

CRQ of word-unit reading times

As can be seen in Figure 8, all measures drop reliably for both reader groups with repeated reading of word units: %RECurrence, $F(1,108) = 13.61$, $p < .001$, %DETerminism, $F(1,108) = 24.84$, $p < .001$, and TrappingTime, $F(1,108) = 16.69$, $p < .001$. Also in the word-unit condition, we observe a reliable interaction between repeated readings and reader group for TT and while TT decreases for both graduate readers and undergraduate readers, the difference is much more pronounced for undergraduate readers, $F(1, 108) = 5.02$, $p < .05$. These precipitous drops in the CRQ measures all pertain to the constraints available that might support shared dynamics among readers. Apparently, sources of constraint available in the first reading of the text, in the word unit condition, have a reduced effect in the subsequent reading.

The decreases in shared %REC and shared %DET indicate that less shared and more idiosyncratic dynamics emerged with a subsequent word-by-word reading. The decrease in TT also indicates that reading times are less often in the same neighborhood, from one word to the next of the story text, and less often still for undergraduates compared to Ph.D. students. This result is opposite in trend from what the individual RQAs revealed in %DET and TT, so the results cannot be due to less determinism or structure at the level of the individual reader.

Re-reading a story word-by-word could disperse trajectories in a reader's reconstructed phase-space and reduce the likelihood that readers share trajectories of reading times. This would decrease %REC, %DET, and TT in the CRQ variables of the re-reading. In other words, one source of the increasingly idiosyncratic performances could stem from data that are more randomly dispersed due to extraneous sources of random noise and this possibility fits the outcomes of the CRQ. But it is not supported by other data.

For instance, the standard deviations of reading times are apparently smaller for repeated readings, word-by-word, suggesting a more stable performance and a reduction of extraneous noise. Also the monofractal analyses discovered a less random pattern of noise in the repeated reading, together with a narrower spectrum of multifractal dynamics. And the RQA revealed trends toward more determinism in re-reading. These results together rule out extraneous sources of random noise as the reason for less shared structure in reading times across participants (cf. Riley & Turvey, 2002).

Possibly, these results imply a more flexible coupling between readers and the self-paced reading task with repeated reading (cf. Van Orden et al., 2011). We may speculate that participants eventually discover the tapping strategy in all word-by-word presentation conditions. This strategy hypothesis is most strongly supported by the high rates of %DET in both individual RQAs and joint CRQ analyses, and the fact that similar results were found for both English literature, Ph.D. students and undergraduate readers. In the first reading, however, the text supplies perturbations to this strategy in rare words, or difficult passages, or in other aspects that compose unsystematic perturbations of reading performance. Re-reading the same text word-by-word reduces these sources of perturbation, which are sources of random noise in individual performance.

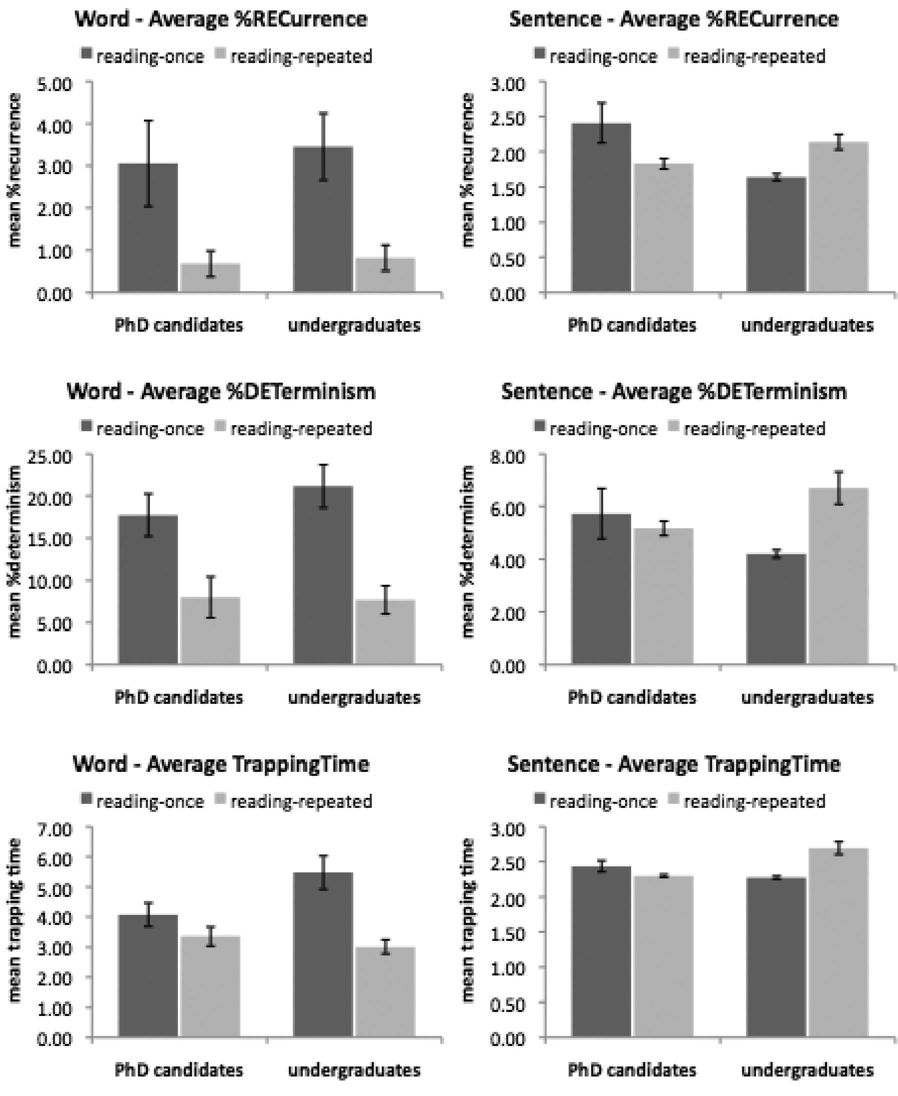

Figure 8. Cross recurrence analysis of word and sentence reading times with standard error brackets.

These same sources of shared perturbation are also sources of shared dynamical structure in comparisons between individuals, bolstering the CRQ variables in the first reading. A re-reading of the story (word-by-word) will benefit from the reduction of perturbations, as indexed by reduced standard deviations, less white noise structure, a narrower multifractal spectrum, and trends of increasing RQA variables. However, the re-reading also reduces the shared dynamical structure that rare words, difficult passages, and other surprises provide to readers. And it is the reduction of these sources of shared structure that reduce the CRQ variables.

CRQ of sentence unit reading times

The sentence unit condition yielded reliable 2-way interactions between reader group and number of readings for all three CRQ variables: %RECurrence, $F(1,108)=11.62$, $p<.001$, %DETeterminism, $F(1,108)=6.71$, $p<.05$, and TrappingTime, $F(1,108)=20.16$, $p<.001$. %REC, %DET, and TT appear to stay about the same, relatively, for graduate readers, whether reading the story for the first time or re-reading the story. In contrast, %REC, %DET, and TT increase reliably when undergraduate readers re-read the same story. This outcome would be expected if graduate students' reading performance is always close to a performance ceiling, even reading a new story, sentence-by-sentence, for the first time. That is, graduate students in English literature are über-fluent, compared to fluent undergraduate readers.

The hypothesis that PhD. students in English literature are über-fluent is patently intuitive and is supported across the board by the total pattern of outcomes. Throughout all the analyses of sentence-by-sentence reading, graduate students produced more similar scores on all variables between first and repeated readings, when compared to undergraduates. For both RQA and CRQ variables, graduates student scores never differ reliably in the statistical analyses and the apparent changes are always in the direction opposite to that of undergraduate readers. Undergraduate's RQA and CRQ scores always increase, often reliably, after the first reading. This would be the case if the undergraduates' sentence reading times are more uniform in the repeated reading, which may result from a reduction of text-based perturbations affecting the first reading.

The results of sentence unit reading times make sense if the graduate students are near the performance ceiling of reading fluency, which we can only see clearly when first and subsequent readings are compared and the presentation units are right (sentence units). Wallot and Van Orden (2011) found a similar outcome using phrase-unit presentations. Reading times to sentence units are rate limiting compared to spacebar pressing, which is not true of reading times to word units, and so sentences allow the more telling portrait of reading performance. Also, the performance of undergraduate readers became more stable and deterministic when rereading the story sentence-by-sentence. They repeated more of the same reading times at the same locations in the phase space of their performance dynamics.

Conclusions

Both conventional and nonlinear analyses reveal aspects of self-paced reading performance, but they do not reveal the same aspects either in kind or magnitude. Also, all else equal, the results of the nonlinear analysis must always trump analyses that derive from the General Linear Model, in the sense that the nonlinear results will apply more generally than the more narrowly conceived linear analyses (analyses that assume continuity and proportionality, see Mitchell, 1999).

For example, the monofractal analysis identified the patterns of variation across individual reader's reading times as fractional Gaussian noise and fractional Gaussian noise has indefinite variance. For human event time data this means that the more numerous the sample of data collected, the more extensively a project will explore ranges of values that include rare and very large (slow) event times, due to the pronounced skew toward slow times in human event times.

When data sets do not fit the assumptions of the General Linear Model, many scientists resort to the assertion that ANOVA (for example) is robust in the face of heterogeneous variability — sometimes as a last resort, it seems. Nonlinear analysis techniques open up new options to those who investigate complex data sets or happen to have collected data that do not fit the assumptions of the General Linear Model well.

The self-paced reading times exhibited pronounced heterogeneous variability (see Figure 3). The data have a defined population mean, in principle, but in practice it may be impossible to estimate this mean reliably. Indefinite heterogeneous variance implies sample variances that are also indefinite and unstable. The sample variance also impacts the empirical estimate of the mean of repeated measures taken from a single participant, like the data from the self-paced reading task. Differences between means remain meaningful, nonetheless, although what they mean must always be interpreted within the prior context of the results that come from nonlinear analyses. This is because the nonlinear analyses are more general and the linear analyses are the special cases.

Nonlinear methods of analysis may also trump linear methods by revealing differences in kind as well as in magnitude. For instance, the conventional analyses of self-paced reading data could easily have led us astray without the corrective force of the nonlinear outcomes. The conventional analyses suggested to us that über-fluency has a causal basis in — is "localized" in — sentence knowledge, which turned out to be wrong-headed. The conventional analysis had tricked us into asking: What was it about sentences versus words that distinguishes über-fluency from ordinary fluency?

In actual fact, the participants were doing one kind of task in the word-unit condition and a different kind of reading task in the sentence-unit condition. And it was the nonlinear analyses that clarified how and where we were being misled. Thus the better question turned out to be: What strategy distinguishes word-by-word presentations from sentence-by-sentence presentations?

With an answer to that question we could refute the false promise of the conventional results, and the answer was got from nonlinear analyses of reading plus previously published nonlinear analyses of tapping and simple reaction time. How much longer would it have taken to recover from the error, without the help of nonlinear methods? Consider that word-by-word presentation conditions remain a standard tool of self-paced reading methods with college student participants, and have remained so, unquestioned, for decades upon decades of reading research.

In the end, it was sentence-by-sentence reading conditions that distinguished über-fluent Ph.D. students in English literature from fluent undergraduate readers. And the fleshed out nature of the distinction came almost exclusively from gains to be had by fluent undergraduate readers in a repeated reading of the same story. Those gains were not in superficial aspects of performance such as reading speed, however, they were to be had in the more subtle aspects that refer to stability, order, and determinism.

Acknowledgements

Preparation of this article was supported by NSF grants to Guy Van Orden (BCS #0642716; BCS #0843133; DHB #0728743).

References

Bassingthwaighte, J. B., Liebovitch, L. S., & West, B. J. (1994). *Fractal physiology*. New York: Oxford University Press.

Brown, C., & Liebovitch, L. (2010). *Fractal analysis*. Los Angeles: Sage.

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.

Carello, C., & Moreno, M. (2005). Why nonlinear methods? In M. A. Riley, G. C. Van Orden (eds.), *Contemporary nonlinear methods for behavioral scientists: A webbook tutorial* (pp. 1–25). Retrieved August 23, 2009, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep

Castillo, R. D., Van Orden, G., & Kloos, H. (in press). The embodiment of time estimation. In A. Avatakis (Ed.), *Multidisciplinary Aspects of Time and Time Perception*.

Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long term memory processes ($1/f^{\alpha}$ type) in human coordination. *Physics Review Letters*, *79*, 4501–4504.

DeGrado, L. P. (2003). *The Arelis complex*. Retrieved August 23, 2009, from http://www.authorsden.com/visit/viewshortstory.asp?id=9632

Deligniéres, D., Torre, K., & Lemoine, L. (2009). Long-range correlation in synchronization and syncopation tapping: A linear phase correction model. *PLoS ONE*, *4*, e7822. doi:10.1371/journal.pone.0007822

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, *108*, 33–56.

Gilden, D. L. (2009). Global model analysis of cognitive variability. *Cognitive Science*, *33*, 1–27.

Holden, J. G. (2005). Gauging the fractal dimension of response times from cognitive tasks. In M. A. Riley, G. C. Van Orden (eds.), *Contemporary nonlinear methods for behavioral scientists: A webbook tutorial* (pp. 267–318). Retrieved August 23, 2009, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep

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*. doi: 10.1037/a0020991.

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

Ihlen, E. A., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: beyond $1/f^{\alpha}$ fluctuation. *Journal of Experimental Psychology: General*, *139*, 436–463.

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.

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

Marwan, N. (2011). How to avoid potential pitfalls in recurrence plot based data analysis? *International Journal of Bifurcation and Chaos 21,* 1003–1017.

McAuley, J. D., Jones, M. R., Holub S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Lifespan development of timing and event tracking. *Journal of Experimental Psychology: General*, *135*, 348–367.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge, UK: Cambridge University Press.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (ed.), *Visual Information Processing* (pp. 283–308). New York: Academic Press.

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

Rayner, K., & Pollatsek, A. (1994). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Riley, M. A., & Turvey, M. T. (2002). Variability and determinism in motor behavior. *Journal of Motor Behavior*, *34*, 99–125.

Seuront, L. (2010). *Fractals and multifractals in ecology and aquatic science*. Boca Raton, FL: CRC Press.

Shockley, K. (2005). Cross recurrence quantification of interpersonal postural activity. In M. A. Riley, G. C. Van Orden (eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26–96). Retrieved August 23, 2009, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep

Shockley, K., Butwill, M., Zbilut, J., & Webber, C., (2002). Cross recurrence quantification of coupled oscillators. *Physics Letters A*, *305*, 59–69.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand, L.-S. Young (eds.), *Lecture Notes in Mathematics*, *898*, 366–381.

Van Orden, G. (2010). Voluntary performance. *Medicina*, *46*, 681–594.

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

Van Orden, G., Kello, C. T., & Holden, J. G. (2010). Situated behavior and the place of measurement in psychological theory. *Ecological Psychology*, *22*, 24–43.

Van Orden, G., Kloos, H., & Wallot, S. (2011). Living in the pink: Intentionality, Wellbeing, and Complexity. In C. A. Hooker (Ed.), Philosophy of Complex Systems (pp. 639–684). Handbook of the Philosophy of Science (Vol. 10). Amsterdam: Elsevier.

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

Wallot, S., & Van Orden, G. (2011). Grounding Language Performance in the Anticipatory Dynamics of the Body. *Ecological Psychology 23*, 1–28.

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 August 23, 2009, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jep

Author's address

S. Wallot or G. Van Orden
CAP Center for Cognition, Action, and Perception
Department of Psychology
University of Cincinnati
Cincinnati, OH 45221-0376
USA

wallotsn@mail.uc.edu
guy.van.orden@uc.edu