

Journal of Experimental Psychology: Human Perception and Performance

Fractal 1/f Dynamics Suggest Entanglement of Measurement and Human Performance

John G. Holden, Inhyun Choi, Polemnia G. Amazeen, and Guy Van Orden

Online First Publication, December 6, 2010. doi: 10.1037/a0020991

CITATION

Holden, J. G., Choi, I., Amazeen, P. G., & Van Orden, G. (2010, December 6). Fractal 1/f Dynamics Suggest Entanglement of Measurement and Human Performance. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0020991

Fractal $1/f$ Dynamics Suggest Entanglement of Measurement and Human Performance

John G. Holden
University of Cincinnati

Inhyun Choi, Polemnia G. Amazeen
Arizona State University

Guy Van Orden
University of Cincinnati

Variability of repeated measurements in human performances exhibits fractal $1/f$ noise. Yet the relative strength of this fractal pattern varies widely across conditions, tasks, and individuals. Four experiments illustrate how subtle details of the conditions of measurement change the fractal patterns observed across task conditions. The results call into question whether measurement noise and measured signal can be distinguished in human performance, suggesting that human performance is inextricably entangled with measurement context. Perhaps, though, a hypothesis of soft assembly of human performance can circumvent the conundrum (e.g., Turvey, 2007).

Keywords: $1/f$ scaling, pink noise, self-organization, cognitive dynamics, interaction dominant dynamics

Do the same thing twice and the second time it will be measurably different, every time. The inherent variety or multiplicity in behavior is a basic fact, and the measurement of variation is the empirical basis by which to understand ordinary behavior. That is why scientists study behavioral changes in laboratories. In actual fact, when a scientist says the word *behavior* that scientist is talking about changes in some particular measurements of behavior. This is why behavioral scientists become experts in statistics to analyze and explain behavioral variability. And really, variation writ large is the thing explained in every theory offered by psychology.

Variation in Natural Fractals

The endogenous variation of human behavior forms a fractal pattern (Gilden, 2001). Relatively clear pictures of endogenous variation can be got by repeatedly measuring the same person in the exact same act, holding all variables constant. In the best

recipe, use a simple task with which to measure human behavior and repeat the measurement trials over and over again, to collect a large sample of hundreds or thousands of data points, as a trial-ordered time series. If almost nothing changes from trial to trial, except another measurement is taken, then the primary source of variation will be variation endogenous to the behavior that is measured (Kello, Anderson, Holden, & Van Orden, 2008).

The fractal pattern of variation is usually portrayed in a spectral analysis of repeated measurements. Repeated measurements can be fully characterized by the size of changes across measured values and the frequency of changes of a particular size—how often changes of a particular size occur. The fractal pattern relates size and frequency of variation in a scaling relation. The scaling relation between the size of changes, and how often changes of that size occur, is inversely proportional on logarithmic scales. As portrayed in Figure 1, the amplitudes of variation can be seen to scale with frequency of variation such that $S(f) \approx f^{-.94}$.

The Y-axis in Figure 1's spectral portrait corresponds to how big the changes are and the X-axis is their frequency of occurrence. Since $f^{-\alpha} = 1/f^\alpha$, the size of change, $S(f)$, is proportional to the frequency, f , of change, as $S(f) \approx 1/f$. The scaling exponent α that describes the relation between amplitude and frequency of variation is depicted as $\alpha = .94$, which derives from the slope of the line or spectral slope in Figure 1. As with all fractals, the ratio that the scaling exponent refers to is invariant across the changes of different size and frequency. In what follows, we report scaling *exponents*.

Natural fractal patterns in time exhibit $1/f$ scaling, but spectral exponents equal to one are not usually observed. A notable exception has been healthy human heart-rate variability, which often displays a spectral exponent very near to one (e.g., Eke, Hermán, Kocsis, & Kozak, 2002), but many established examples of $1/f$ scaling yield exponents between zero and one or between one and two. Mandelbrot and Wallis (1969b/2002) describe a large sample of yearly tree rings yielding scaling exponents that average about $\alpha = .43$. Annual precipitation statistics have average exponents of about $\alpha = .48$, and

John G. Holden and Guy Van Orden, CAP, Center for Cognition, Action & Perception, Department of Psychology, University of Cincinnati; Inhyun Choi and Polemnia G. Amazeen, Department of Psychology, Arizona State University.

We acknowledge financial support from the National Science Foundation Grant BCS-0446813 to John G. Holden, BCS-0447039 to Polemnia G. Amazeen, BCS-0642718 to John G. Holden and Guy Van Orden, and BCS-0642716, BCS-0843133, and DHB-0728743 to Guy Van Orden. Inyun Choi wrote the data collection software and conducted Experiments 1 and 2 as part of a First-Year graduate research project. We thank Adriana Aldana, Gerardo Ramirez, Karina Shokat-Fadai, and Gregory Stewart for their dedicated assistance during the data collection phase of this project.

Correspondence concerning this article should be addressed to John G. Holden, Center for Cognition, Action, & Perception, Psychology Department, University of Cincinnati, PO Box 210376, Cincinnati, OH 45221-0376. E-mail: john.holden@uc.edu

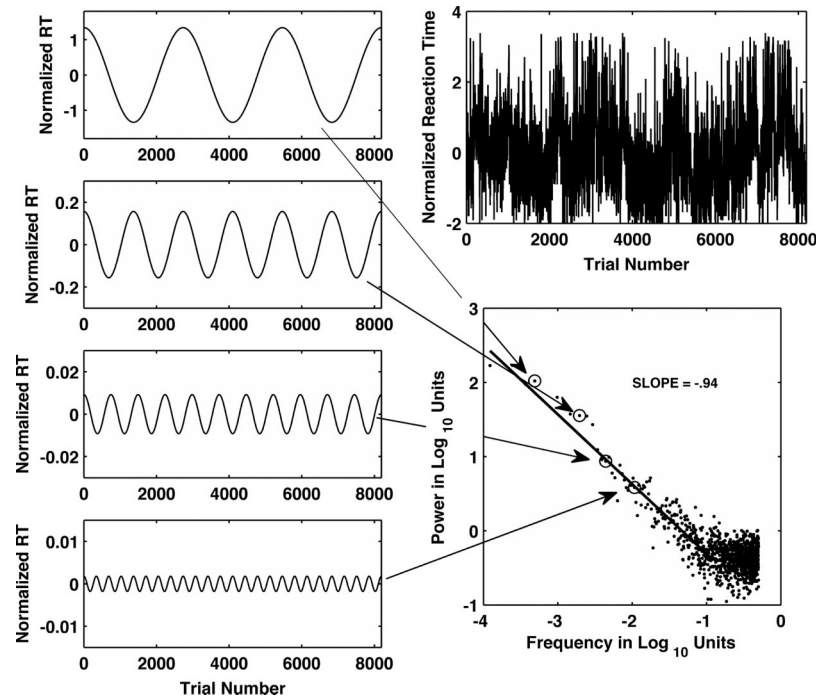


Figure 1. An example of $1/f$ scaling across 8192 reaction times from Van Orden, Holden, & Turvey, (2005) is depicted. The top right plot in the Figure is a graph of 8192 normalized simple reaction times, graphed according to the trial order in which they were collected. The X-axis portrays the trial number, and the Y-axis is reaction time, in normalized units. Directly below the graph of reaction time data is a plot of the spectral coefficients of the same reaction time data: the output of a power spectral density analysis. A spectral analysis approximates the complex pattern of oscillation entailed in simple reaction times by decomposing it into a set of basic sinusoidal oscillations. Conceptually, the procedure resembles a prism decomposing white light into its elemental frequencies. Each point on the X-axis of the spectral plot depicts a specific frequency of oscillation of a sine/cosine wave used in the approximation; one of a large number of sine waves ordered from lowest to highest frequency along the X-axis. The Y-axis of the spectral plot depicts the relative power or magnitude of the oscillations of each frequency. Both axes represent logarithmic scales, which means that the extent of sine wave frequencies and their relative amplitudes spans a wide range. The four plots on the left side of the Figure illustrate particular sine waves that are depicted as points in the spectral plot. The top plot depicts a very low-frequency sine wave, one of the three lowest frequencies required to approximate the graph of simple reaction times. The specific point that each wave is represented by is surrounded by an open circle, and indicated by the arrow that extends from the sine-wave plot to its location in the spectral plot. It is important to note that the successive axes of the sine plots had to be enlarged to make the oscillations visible. Also, the example sine waves all use the same phase, or starting point. In reality, the point in the cycle that creates the best match with the data is used as the origin for each particular sine function. If the sine waves corresponding to each frequency were generated, with the appropriate phase, and were then all added together, the resulting signal would essentially regenerate the original trial-series signal. This figure was reprinted from "Situating Behavior and the Place of Measurement in Psychological Theory," by G. C. Van Orden, C. T. Kello, and J. G. Holden, 2010, *Ecological Psychology*, 22, p. 24–43 with permission of the publisher, Taylor & Francis Ltd.

the classic yearly series of minimum levels of the Nile river yields a scaling exponent of about $\alpha = .82$, whereas the scaling exponent of the Nile's yearly maximums is $\alpha = .68$ (transformed from Hurst exponents). Long-range correlated chemical structures of DNA sequences yield average scaling exponents of $\alpha = .82$ for viruses, $\alpha = .86$ for plants, and $\alpha = .84$ for mammals, all appreciably different from $\alpha = 1$ (Voss, 1992).

Variation and Task Demands

Like other natural fractals, human performance yields a variety of scaling exponents. In a well-controlled recipe of task demands, the

variation across repeated measurements forms a fractal pattern as we have noted—never a perfect or ideal fractal pattern, but one that appears to be colored to various degrees by random white noise (Thornton & Gilden, 2005). The subtle changes in task demands, the way behavior is measured, change the slope of the spectral plot and the corresponding estimate of the scaling exponent. Yet this scaling exponent indexes the kind of variation and therefore the kind of behavior that is observed. Variable scaling exponents require explanation.

One possibility is that variation due to task demands is simply layered on top of the fractal pattern, changing the overall empirical

pattern of variation in the bargain. If so, then the simplified schematic in Figure 2 introduces a prediction that can be tested. The prediction concerns a manipulation of random white noise injected into each trial to perturb performance, to add white noise to human performance. White noise has a scaling exponent of zero, which is a flat line parallel to the X-axis in the figure. The amplitude of white noise is its intersection point on the Y-axis. As the amplitude of the added white noise is increased, it will encroach upon the sloped line of the spectral plot from lower to higher amplitude, and higher to lower frequency.

Low amplitude white noise will have a significant impact on the low-amplitude high-frequency region of the spectral plot, whereas sufficiently high amplitude white noise will impact both the high-frequency and the low-frequency regions of the spectral plot. This differential effect of white noises of different amplitudes on the high-frequency versus low-frequency regions of the plot generates a prediction. It predicts an interaction effect between the amplitude of noise and whether scaling exponents were calculated to reflect the overall spectrum or to emphasize the higher-amplitude lower-frequency region of the spectrum.

For example, Figure 3 depicts two regression lines, overlaid on a power spectrum typical of present data, and the regression line through the coefficients of the overall spectrum has shallower slope and a smaller scaling exponent, compared to the regression line through the 25% of the coefficients at the lowest frequencies and highest amplitudes. A simulated portrait of the prediction can be got from Figure 4. This figure portrays scaling exponents from multiple simulations that layer different amplitudes of white noise on top of an artificial $1/f$ fractal signal, $\alpha = 1$.

In Figure 4, amplitude of noise increases along the X-axis. The uppermost line graph portrays average scaling exponents estimated

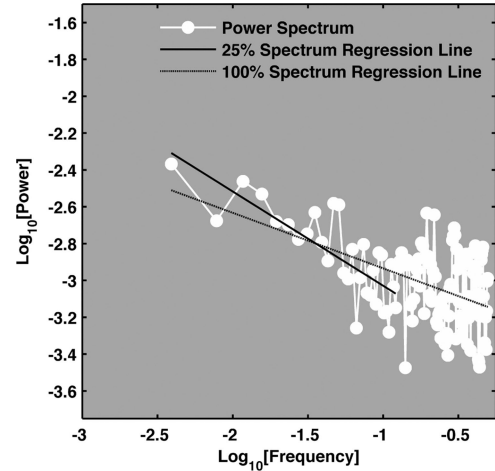


Figure 3. The outcome of a 127-point (Welch) power spectrum, based on 1024 response times is depicted as is depicted by filled white markers (see Holden, 2005 for details). The X-axis represents frequency on a logarithmic scale, likewise, the Y-axis represents power on a log scale. The regression line resulting from using the lowest 25% of the frequency-power pairs is represented by the solid line. The dashed line represents the regression line that results from using 100% of the frequency-power pairs in the regression. The power spectrum is relatively level in the high-frequency range (e.g., log frequency > -1). This biases the regression line towards a shallower value, and is the basis for the differences between the scaling exponents derived using the 25% and 100% Spectrum in this article.

using calculations of a 25% Spectra, estimating scaling exponents using the highest-amplitude lowest-frequency region of the power spectrum. This 25% of the spectrum is less affected by relatively low amplitude white noise. The lowermost line graph in Figure 4 portrays the average of scaling exponents estimated from 100% Spectra, estimating scaling exponents using the full spectrum of amplitudes and frequencies, high and low. The overall prediction comes from the overall pattern of change with increasing amplitudes of white noise: scaling exponents all close to $\alpha = 1$ when white noise is eliminated, that diverge with intermediate amplitudes of white noise, only to converge again toward $\alpha = 0$ with white noise of sufficient amplitude.

The differential effect of white noise on scaling exponents from 25% and 100% Spectra predicts an interaction effect between noise amplitude and 25% versus 100% Spectra calculations. The model that inspired the prediction, itself combines white noise with a fractal pattern to approximate the changes in variation induced by exogenous task demands (Gilden, 2001; Thornton & Gilden, 2005). Endogenous fractal-like patterns have also been simulated using sums of sine waves plus random noise or similarly by autoregressive-moving average (ARMA) systems that decay over a range of time scales (Farrell, Wagenmakers, & Ratcliff, 2006a; Wagenmakers, Farrell, & Ratcliff, 2005). In all these simulations different amplitudes of white noise are layered on top of the sine waves (Ward, 2002). Thus they all predict the interaction effect portrayed in Figure 4.

The experiments we report test this common prediction, modulating the exogenous variation introduced by task demands in the amplitude of injected white noise. Next, we describe four experi-

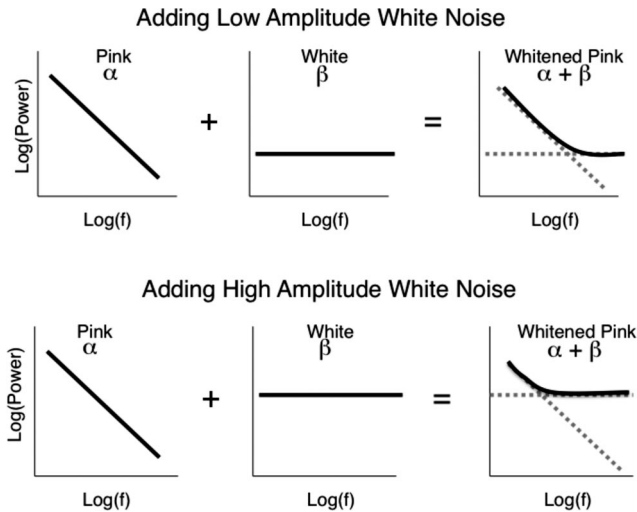


Figure 2. This schematic describes the essential features of the Thornton and Gilden (2005) model. The upper three graphs supply a schematic depiction of the impact on the power spectrum of adding a low-amplitude, uncorrelated white noise to an ideal $1/f$ noise. The lower three graphs supply a schematic depiction of the impact on the power spectrum of adding a high-amplitude, uncorrelated white noise to an ideal $1/f$ noise. The alpha (α) parameter corresponds to the scaling exponent described in Figure 1, the beta (β) parameter corresponds to the SD of the white noise, relative to the pink, or $1/f$ noise.

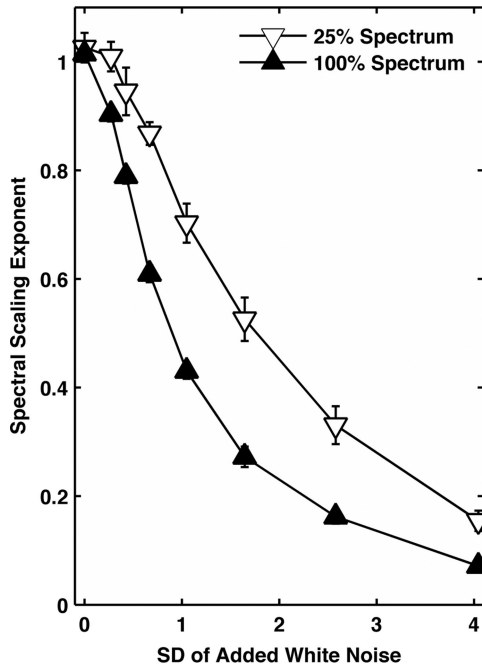


Figure 4. The X-Axis indexes the relative standard deviation of synthetic, uncorrelated, white noise, added to a synthetic $1/f$ noise (i.e., $\alpha = 1$), such that $SD = 1$ for the $1/f$ signal. The Y-Axis depicts the value of the scaling exponents returned, computing them from either the lowest 25% of the spectral coefficients (25% Spectrum) or by using 100% of the spectral coefficients (100% Spectrum). Absent any added source of white noise ($SD = 0$), both estimation methods return identical scaling exponents. However, elsewhere, scaling exponents from the 100% Spectrum are affected more dramatically than those from the 25% Spectrum until both exponents approach $\alpha = 0$ of white noise in the case of a very large SD of added uncorrelated variability. The 100% Spectrum is differentially sensitive to whitening that affects the low amplitude, higher frequency estimates of a spectrum because the synthetic (or ideal) $1/f$ noise, itself, has a low amplitude in the high-frequency range. SD = standard deviation.

ments to examine how exogenous variation changes the fractal pattern. Each experiment includes a manipulation of the amplitude of uncorrelated white noise, the amplitude of exogenous *whitening*, injected into the temporal coupling between the participant and the trial series of the laboratory task. White noise is injected into downtimes of the trial sequence, so to speak, within each experimental trial.

That is, white noise of different amplitude is injected between a “ready signal” and a “signal to respond,” which control the durations of inter-stimulus intervals (ISI) in a simple reaction time task. In other experiments, white noise of different amplitudes is injected into downtimes between trials, which control the duration of inter-trial intervals (ITI). As we report, the different amplitudes of injected white noise all impact the 25% Spectra and the 100% Spectra to the same degree, failing to confirm the predicted interaction effect, and failing to confirm the hypothesis of layered white noise on top of a $1/f$ signal.

We confess up front, before wading through the details of the experiments, that we fully expected the layering hypothesis to fail. We derived the layering hypothesis from thinking about entanglement of signal and noise and how that idea might be manifest (c.f.

Van Orden, Kello, & Holden, 2010). In the contrast with layering, the idea of entanglement lead us to the present experiments, fully aware that we might bring into question a time-honored and fundamental distinction of experimental psychology. What might it mean if human competence is equivalent to human performance and the latter is fully entangled with laboratory contexts and procedures?

Whitening Injected After a Ready Signal and Before a Signal to Respond

Psychologists have long known that reaction time performance is sensitive to changes in the “foreperiod” (the time between a ready signal and a signal to respond). However this effect was thought to be a local effect specific to each particular trial measurement (e.g. Woodworth & Schlosberg, 1954). The first few experiments all test this assumption in simple reaction time tasks that present a warning or ready signal followed by a signal to respond. Each participant responded identically in each trial with a button press response, in one response condition, or a vocal response in the other (cf. Gilden, Thornton, & Mallon, 1995; Van Orden, Holden, & Turvey, 2003; Wagenmakers, Farrell, & Ratcliff, 2004). This first experiment injected five different amplitudes of uncorrelated white noise, *whitening*, into the duration between each trial’s ready signal and the signal to respond. The levels of the manipulation differed in whether the time intervals varied not at all, not much, a bit more, more than that, or widely in a rank order of amplitude.

Method

Participants. Fifty students participated to fulfill a course requirement of an introductory psychology course at Arizona State University. Ten undergraduate students were assigned randomly to each of five conditions. Each student participated in both a manual and a vocal response condition. Presentation order of the response conditions was counterbalanced across participants and conditions.

Procedure. There were five levels of ISI variability, we first define a baseline condition, and then describe the four remaining conditions. Each trial began with a ready signal (+++) displayed in the center of a computer monitor. The ready signal was visible for 171 ms and then replaced by a blank screen. The blank screen was visible for 700 ms and then replaced by a signal to respond (#####, displayed in the center of the computer monitor). The signal to respond remained visible until a response was recorded, or a maximally 5004 ms. Participants responded by pressing a joystick button or saying /ta/ into a microphone, in the vocal response condition. Time from each response until the next ready signal was 418 ms. The procedures to this point define a relative baseline condition, similar to conditions in previous reports of reliable fractal patterns of $1/f$ scaling.

The baseline condition used a 700 ms ISI duration on every trial for the blank screen between the ready signal and the signal to respond. The four other conditions varied the presentation duration of the blank screen across trials, choosing uniformly and randomly sampled durations. The mean duration was always 700 ms but the width of dispersion was manipulated in the standard deviation of durations of the blank screen. The narrowest dispersion of five durations was 676 ms, 688 ms, 700 ms, 712 ms, and 724 ms ($M =$

700, $SD = 18.97$ ms). The widest dispersion was 512 ms, 606 ms, 700 ms, 794 ms, and 888 ms ($M = 700$, $SD = 148.63$ ms). The remaining conditions were likewise evenly spaced across five durations such that the $M = 700$, $SD = 37.32$ in one condition and $M = 700$, $SD = 74.31$ in the other. With the baseline condition, in which $M = 700$, $SD = 0$, this yielded five conditions with SD s of: 0, 18.97, 37.32, 74.31, and 148.63 ms, respectively, which we will call the 0, 19, 37, 74, and 149 conditions.

Participants were instructed to respond immediately after the signal to respond (i.e., #####) by pressing a joystick trigger-button or saying /ta/ into a microphone. Response times were measured from the onset of the signal. Each session consisted of 64 introductory trials followed by 1100 measurement trials. Introductory and measurement trials were otherwise identical. Participants completed the 64 introductory trials within about 3 min and the 1100 trials within about 20 min. A 5-min break separated the manual and vocal response sessions. Participants usually completed the two sessions within 60 min.

Apparatus. A standard PC controlled signal presentations and data collection. The perturbation conditions varied in the range of durations of the blank-screen which controlled the time from the offset of the ready signal to the onset of a signal to respond. A video-controller refresh-cycle freezing routine allowed the blank screen durations to be controlled to the nearest millisecond (Buhrer, Sparrer, & Weitkunat, 1987) and the blank screen durations were varied in even multiples of the 11.76 ms vertical raster-refresh cycle of a 85 Hz CRT video display. The ready signal and signal to respond were presented in the center of the video monitor. A 6 ms allowance was incorporated into the reported display durations, the time required for the raster to pass the center of the display during its refresh cycle.

In the manual condition, participants responded by pressing the primary trigger button of a standard four-button joystick with their index finger. The direct input (non-buffered) joystick was positioned directly in front of the participant's dominant hand, and accurately recorded response time to the nearest ms. A headset microphone detected vocal responses. The microphone was omnidirectional and sensitive to frequencies ranging from 20 Hz to 20,000 Hz. The distance between the microphone and participants' mouth was approximately 3.5 cm. The microphone sensitivity was adjusted to detect a voice volume greater than 11 dB. The level of surrounding noise was approximately 4 to 5 dB.

Results

We report the average scaling exponents of participants in each condition, derived using 25% Spectra and 100% Spectra power spectrum analyses. Power spectrum analysis is the typical analysis to examine the fractal patterns of variation in the behavioral literature, and in other disciplines. Also, the spectral approach is sensitive to the effects of the exogenous perturbations that we manipulate, which would become a weakness if we were seeking to minimize or eliminate contamination of exogenous factors (see Caccia, Percival, Cannon, Raymond, & Bassingthwaite, 1997; Eke et al., 2000; 2002).

Spectral analysis requires pre-preparation of the data. First, observations greater than 1000 ms were eliminated from each series. Second, response times that fell beyond ± 3 SD s from the series mean were eliminated. In every case, more than 1024

observations remained following the trimming. Third, we truncated the initial observations in each series so that 1024 observations remained in the series. Finally, each trial series was normalized. We followed these same treatments of data throughout this project, noting any exceptions.

Two spectral analyses were conducted for each data series. We call the first method the 25% Spectrum because this analysis calculates scaling exponents using the 25% of spectral coefficients that correspond to the lowest frequency and highest amplitude variation captured in a spectral analysis. These lowest-frequency 25% of coefficients were alone included to estimate the slope of the power spectrum and then the scaling exponent (compare Figure 3).

Hypothetically, 25% Spectra target embodied processes on the slowest time scales that spectral analyses can capture reliably, and least perturbed by a layer of exogenous noise. We expected the amplitude of white noise injected into ISIs to be inversely related to average scaling exponents in all cases. For the vocal response data, noise amplitude accounted for 23% of variability in scaling exponents of 25% Spectra, $r^2 = .23$, $F(1, 48) = 14.45$, $p < .05$, and 15% of variability for the manual key-press response data, $r^2 = .15$, $F(1, 48) = 8.60$, $p < .05$. The 100% Spectrum incorporating the entire power spectrum produced scaling exponents more perturbed by a layer of exogenous noise. For the vocal response data, noise amplitude accounted for 41% of variability in scaling exponents from 100% Spectra, $r^2 = .41$, $F(1, 48) = 33.80$, $p < .05$, and 25% of variability for the manual key-press response data, $r^2 = .25$, $F(1, 48) = 16.16$, $p < .05$.

If exogenous noise is a layer of noise added on top of endogenous variation, the average scaling exponents of the 25% and 100% Spectra should be further apart at intermediate amplitudes of exogenous noise. This prediction was tested in a 2 (25% vs 100% Spectra scaling exponents) \times 5 (noise amplitudes) mixed design analysis of variance (ANOVA), but no reliable interaction effect is apparent, either in the vocal response data of Figure 5, $F(4, 45) < 1$, $\eta^2 = .01$ or in the manual key-press response data of Figure 6, $F(4, 45) < 1$, $\eta^2 = .02$. The results of this experiment fail to support the layering hypothesis. Next we describe an experiment that injects the same manipulation of noise amplitude into inter-trial-intervals.

Whitening Injected After a Response and Before the Ready Signal

We next inserted the manipulation of the amplitude of white noise into the downtime between each trial response and the next ready signal indicating the beginning of the next trial, into the inter-trial interval (ITI). Now ITI becomes unpredictable while a reliable, constant, predictable ISI controls the duration between the ready signal and the signal to respond. If the effect of injecting white noise previously in the ISI, was simply to make the signal to respond unpredictable, then this problem would be rectified as the white noise manipulation is moved to the ITI.

Method

Participants. Fifty additional introductory psychology students at Arizona State University participated in exchange for course credit. Ten undergraduate students were assigned randomly to each of five conditions. Each student participated in both a

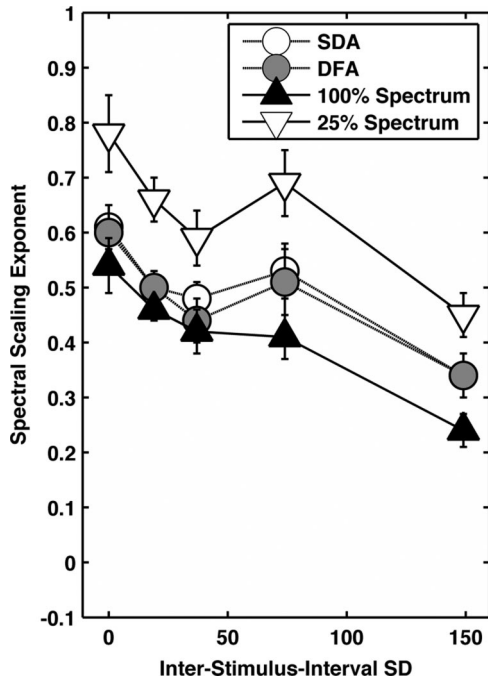


Figure 5. The outcome of the inter-stimulus interval standard deviation manipulation (ISI SD) for vocal responses. The X-Axis indexes the ISI SDs that were introduced between the offset of the fixation stimulus and the onset the target stimulus. The mean ISI was always 700 ms. The white triangles represent the average scaling exponents returned by 25% spectrum estimation method. The filled triangles depict the average scaling exponents returned by the 100% spectrum estimation method. The whiskers indicate standard errors of the mean. A visual examination of the plot might suggest unexpected higher-order quadratic or cubic trends across the average spectral exponents, as the SD of the injected white noise increases. However, the apparent trends were not statistically reliable in hierarchical regression analyses, conducted on the data sets in each of the four experiments. Only the visible linear trends were statistically reliable. Statistical tests for an interaction revealed the difference in the average scaling exponent returned by the two methods was constant across the range of SDs used in the experiment. The white circular symbols depict the average scaling exponents, computed using a *Standardized Dispersion Analysis* (SDA). The filled, gray circular markers depict the mean scaling exponent returned using *Detrended Fluctuation Analysis* (DFA, Bassingthwaighe, Leibovitch, & West, 1994; Holden, 2005; Peng, Havlin, Stanley, & Goldberger, 1995). These methods draw on a mathematical approach distinct from spectral methods. In particular, they do not put as much weight on the higher frequency region of the fractal pattern and thus, like the 25% spectrum, are more impervious to whitening in the higher frequency range of variation (Eke et al., 2002). The scaling exponents of vocal response data computed using fractal analyses replicate the inverse relation with amplitude of ISI white noise, for the SDA, $r^2 = .30$, $F(1, 48) = 20.92$, $r^2 = .23$, $p < .05$, for the DFA, $F(1, 48) = 14.58$, $p < .05$. Thus, the fractal analyses corroborate the outcome of the spectral analyses. DFA = detrended fluctuation analysis; SDA = standardized dispersion analysis; ISI = inter-stimulus interval; SD = standard deviation.

manual and a vocal response condition. Presentation order of the response conditions was counterbalanced across participants and conditions.

Apparatus and procedure. The laboratory apparatus and procedures were identical to those described for the previous case

study except that the manipulation of temporal intervals was injected this time after a response and prior to the next ready signal. The temporal interval between the ready signal and the signal to respond, the ISI, was fixed at 418 ms.

Results

We again report the average scaling exponents of participants in each condition, derived using 25% and 100% Spectral analyses. For vocal response data, noise amplitude was again inversely related to the average scaling exponents in both the 25% and 100% Spectra. The regression analysis of noise amplitude versus the 25% Spectra scaling exponents captured 15% of variability, $r^2 = .15$, $F(1, 48) = 8.70$, $p < .05$, and noise amplitude versus the 100% Spectra scaling exponents captured 24% of variability, $r^2 = .24$, $F(1, 48) = 15.29$, $p < .05$. For the manual key-press response data, noise amplitude captured 32% of variation in scaling exponents

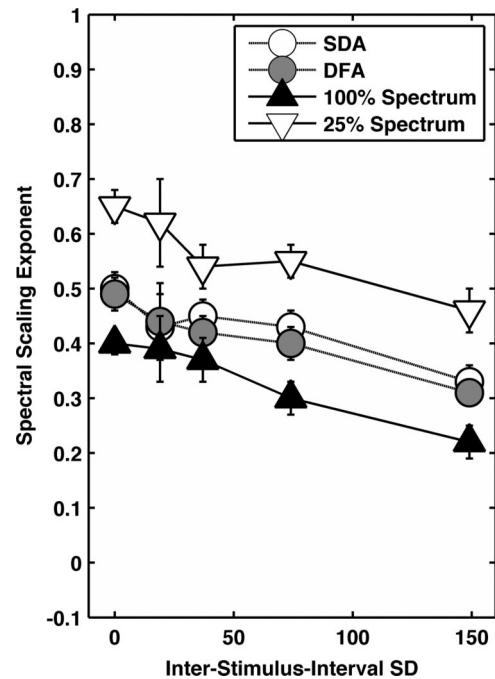


Figure 6. The outcome of the inter-stimulus interval standard deviation manipulation (ISI SD) for manual responses. The X-Axis indexes the ISI SDs that were introduced between the offset of the fixation stimulus and the onset the target stimulus. The mean ISI was always 700 ms. The white triangles represent the average scaling exponents returned by 25% spectrum estimation method. The filled triangles depict the average scaling exponents returned by the 100% spectrum estimation method. The whiskers indicate standard errors of the mean. An interaction test failed to reach significance, indicating the difference between in the average scaling exponent returned by the two methods was constant across the range of ISIs used in the experiment. The white circular symbols depict the average scaling exponents, computed using SDA; the filled, gray circular markers depict the mean scaling exponent returned using DFA. The scaling exponents derived from the fractal analyses yielded a reliable impact of the ISI variability SDA, $r^2 = .18$, $F(1, 48) = 10.62$, DFA $r^2 = .19$, $p < .05$, $F(1, 48) = 11.07$, $p < .05$. DFA = detrended fluctuation analysis; SDA = standardized dispersion analysis; ISI = inter-stimulus interval; SD = standard deviation.

from the 25% Spectra, $r^2 = .32$, $F(1, 48) = 22.69$, $p < .05$, and 26% of the variability from the 100% Spectra, $r^2 = .26$, $F(1, 48) = 17.16$, $p < .05$.

Once again the layering hypothesis was tested in a 2 (25% vs 100% Spectra scaling exponents) \times 5 (noise amplitudes) mixed design ANOVA test for the interaction effect and, just as we observed previously, no statistically reliable interaction effect is apparent in vocal response data, $F(4, 45) < 1$, $\eta^2 = .01$ see Figure 7, nor in the manual key-press response data, $F(4, 45) = 1.904$, $p = .126$, $\eta^2 = .04$ see Figure 8. Thus, the outcome of a second experiment fails to support the layering hypothesis.

Discussion of ISI and ITI Results

The amplitude of white noise changes the overall fractal pattern across a trial series of simple reaction times, whether injected into ISIs or ITIs. The layering hypothesis is not supported because the changes in the fractal pattern are parallel in the 25% and 100% Spectra and proportional to the amplitude of injected noise. These results substantially resolve an early contentious issue in studies of fractal patterns in behavioral variation. Several previous reports contended that fractal patterns are not always or not usually present in human performance. But all these reports overlooked the potential role of a laboratory method to weaken or eliminate fractal

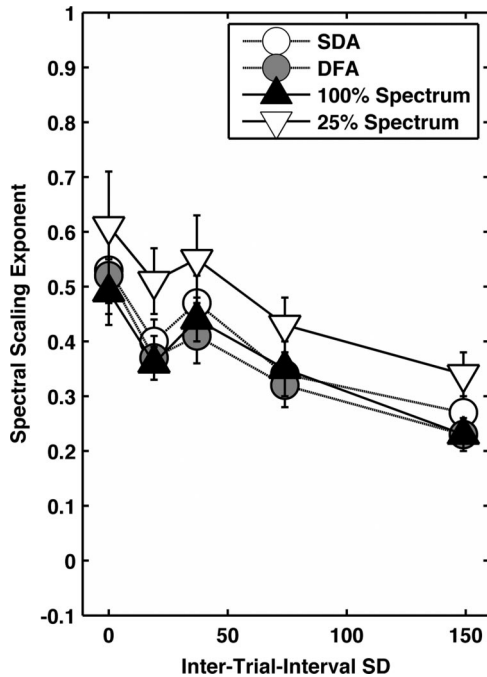


Figure 7. The X-Axis indexes the standard deviation of the inter-trial intervals (ITI) that were introduced between the collection of a response time, and the onset of the fixation stimulus; the mean ITI was always 700 ms. The white triangles represent the average scaling exponents returned by 25% spectrum estimation method. The whiskers indicate standard errors of the mean. The filled triangles depict the average scaling exponents returned by the 100% spectrum estimation method. As for the Vocal Condition, a test for an interaction revealed the difference between in the average scaling exponent returned by the two methods was constant across the range of ITIs used in the experiment.

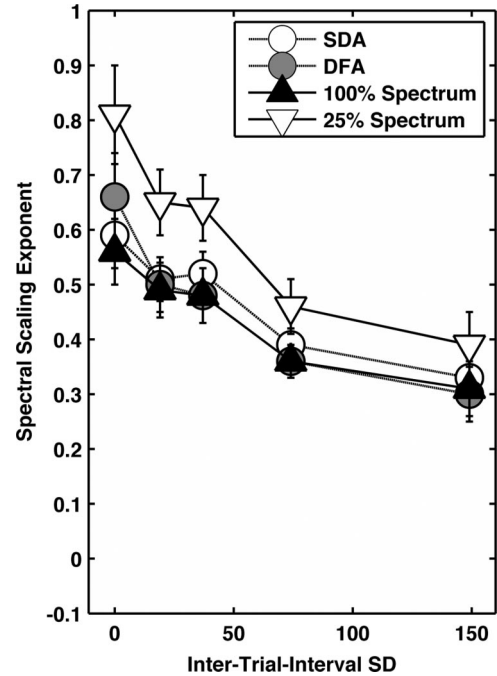


Figure 8. The X-Axis indexes the standard deviation of the inter-trial-intervals (ITI) that were introduced between the collection of a response time, and the onset of a new fixation stimulus, as before the mean ITI was 700 ms. The white triangles represent the average scaling exponents returned by 25% spectrum estimation method. The whiskers indicate standard errors of the mean. The filled triangles depict the average scaling exponents returned by the 100% spectrum estimation method. An interaction test failed to reach significance, indicating the difference between in the average scaling exponent returned by the two methods was constant across the range of ISIs used in the experiment. The white circular symbols depict the average scaling exponents, computed using SDA; the filled, gray circular markers depict the mean scaling exponent returned using DFA. The scaling exponents of manual key-press ITI SD response data computed from fractal analyses captured 32% and 31% of variability, respectively, for SDA $r^2 = .32$, $F(1, 48) = 22.77$, $p < .05$, and for DFA $r^2 = .31$, $F(1, 48) = 21.16$, $p < .05$.

patterns of $1/f$ scaling, while their methods also injected white noise as we did here and in similar tasks (e.g., Gilden et al., 1995; Farrell, Wagenmakers, & Ratcliff, 2006a; Wagenmakers et al., 2004; 2005).

Unsystematic inter-stimulus or inter-trial intervals, randomized conditions or randomized stimuli—or any unsystematic perturbations of method or design—will whiten a fractal pattern that would otherwise be apparent (Kiefer, Riley, Shockley, Villard & Van Orden, 2009; Van Orden et al., 2003). This fact brings into question previous outcomes of classification frameworks to determine the presence of fractal noise (e.g., as implemented by Farrell, Wagenmakers, & Ratcliff, 2006b). The classifier in question computes a difference between log-maximum likelihoods (goodness of fit measures) of a fractal noise description (i.e., Thornton & Gilden, 2005) versus a description that makes no reference to fractal noise (Wagenmakers et al., 2004).

Simply enough, a *positive* difference favors the fractal noise description and a *negative* difference favors a description without

fractal noise. For the present data, the classifier at issue discovered positive differences for smaller amplitudes of injected white noise, decreasing toward zero as the amplitude of the injected white noise increased, and eventually resulting in small negative differences (close to zero) that weakly favor a conclusion that no fractal noise is present (see Table 1). Relying solely on outcomes of the classification framework, we would have demonstrably misrepresented the present data.

In general, changing the circumstances of measurement can reveal robust scaling relations previously deemed absent, or render apparently robust scaling relations more similar to white noise. Yet the means of turning fractal noise into white noise are standard procedures of experimental control—the careful introduction of experimental manipulations or procedures of randomization of conditions or stimuli. If all failures to observe fractal patterns likewise injected white noise through some aspect of method—as in the present studies—then a safe conclusion would be that fractal noises are the default universal character of variation in human performance.

The present results also reinforce our present concern that laboratory protocols change the fundamental structure of variation and thereby behavior itself. If task demands due to injected white noise change the kind of variability in task performance, then scientific methods may not allow researchers to dissociate variation in cognition from variation in the context in which cognition is evaluated. But can we truly rule out the layering hypothesis, in which methods conventionally layer dissociable noise on top of the signal of behavior? We do not wish to rule it out by mistake; that is certain. The question of whether measurement noise is simply added on top of meaningful signals in data is so basic to a science of behavior that it should be asked repeatedly in every reasonable way.

Whitening of Key-Press Response Duration While Sparing Key-Release Duration

Since Helmholtz, response time measurements have been thought to refer to isolated stimulus events, primarily—a response

Table 1
The Sums of the LnL Values Returned for Each SD Condition

All four small <i>SD</i> experiments	ITI & ISI <i>SD</i>				
	0	19	37	74	149
Sum lnL	63.04	22.20	74.83	4.27	−3.52

Note. The sums of the lnL values returned for each *SD* condition. The sums were taken as a function of the *SD* manipulation used in Experiments 1 and 2. Thus we collapsed observations across the two response modes (Manual and Vocal) as well as the interval type (ISI and ITI). Positive sums favor a fractal noise description of the variability; negative sums favor an ARMA description (see Farrell et al., 2006b). The low amplitude *SD* conditions exhibit clear $1/f$ scaling. The 74 and 149 *SD* conditions entail large enough amplitudes of the exogenous variability to whiten the $1/f$ pattern so that it can be modeled by an ARMA description. By way of comparison, Wagenmakers et al. (2004) included a variable ITI (response-stimulus interval or RSI) in each task protocol. The protocols used an average ITI duration of 750 ms in one ITI condition, and 1350 ms in another. Importantly, the standard deviation of the ITI was a constant 115 ms in each task. Sum lnL = Sum of the log-likelihoods. ITI = Inter-trial interval. ISI = inter-stimulus interval.

time measurement estimates the duration of the specific response event that occurs, the response to a presented stimulus for instance. And in the same vein, individual response times were once falsely assumed to be statistically *independent*. Associated properties of the presented stimulus were thought to be driving the participant's response, not aspects of the previously produced responses, for instance (Thornton & Gilden, 2005). In contemporary science, response times are furthermore assumed to estimate the duration of endogenous stimulus processing (however, see Järvilehto, 1998, for an entirely different view).

Keeping foremost in mind the traditional and conventional views at stake, it remains possible to imagine that injected noise merely affects performance superficially, as an artifact of task demands. After all, noise was injected into the ISI and ITI down-times of the trial progression, whereas the crux of the trial exists after the stimulus appears and before the response is made. Perhaps noise may color the estimate of the duration of endogenous processing time but not actually impact the functional or practical basis of executing a simple-reaction-time response.

Functional impacts are demonstrated in combinations of dissociated and associated effects. For instance, if injected noise affects simple-reaction-time responding exclusively, dissociated from other behaviors, then the task demands associated with injected noise must affect the functional basis of simple-reaction-time behavior. Functional dissociations are thought to cleave nature at its functional joints and thus establish the separate and independent functional bases of different behaviors. One source of evidence for a functional dissociation is a lack of correlation between the behavioral measurements of functionally separate and independent behaviors.

The present experiment used injected noise amplitude to dissociate key-pressing behavior from key releasing behavior. Key pressing behavior is indexed by the duration of response time, the time that passes between a signal to respond and contact between the pressed response key and the lead that registers the response. Key releasing behavior is indexed by the duration of time that passes from key contact until the release response breaks the circuit contact. Contact terminates shortly after the contact point as the participant releases the response key to prepare for the next trial.

Repeated measurements of key press times and key release times both exhibit fractal patterns across trials, and their fractal patterns are independent, one from the other (Kello, Beltz, Holden, & Van Orden, 2007). Moreover, uncertainty within a trial about which of two alternative response keys to press whitens the pattern of key press times, compared to no uncertainty. Yet the fractal pattern of key release times, in the same response, is not whitened by the manipulation; the spectral exponent of key release times does not change reliably from one condition to the other. The differential outcomes dissociate the key pressing behavior from the key releasing behavior.

The present experiment replicated and extended this finding injecting noise into the inter-trial interval (ITI), after the key-contact and before the “ready” signal to begin the next trial of the simple reaction time task. If injected white noise whitens scaling exponents of key pressing behavior, but not key releasing behavior, then it dissociates the two behaviors and localizes the effect in functional basis of simple reaction times — according to the

conventions of functional dissociation logic (Van Orden, Pennington, & Stone, 2001).

To insure a sufficiently strong perturbation of ITIs and to extend the range of previous manipulations, the present experiment injected still greater amplitude white noise. A weakness of the present logic, overall, is the unknown relation between magnitudes of empirical noise and the theoretical impact of noise that is at issue. It is possible that the predicted interaction effect of layering noise on top of behavioral signals simply lies outside of the range of our manipulations—so we increase the range of the manipulation.

Method

Participants. Thirty introductory psychology students at California State University Northridge participated in exchange for course credit and a new experimenter performed the study. Ten undergraduate students were assigned randomly to each of three noise amplitude conditions.

Apparatus and procedure. Three additional but larger ITI *SD* conditions were created for the manual response-mode protocol used in previous experiments. The narrowest range of ITI variability was generated by selecting, at random, after each trial, one of five equiprobable inter-trial intervals, ranging in duration from 589 to 811 ms ($M = 700$, $SD = 88$ ms). The intermediate ITI condition presented ITI's ranging from 478 to 922 ms ($M = 700$, $SD = 176$ ms) and the widest ranging ITIs were one of five equiprobable intervals spanning the range from 256 to 1144 ms ($M = 700$, $SD = 351$ ms). This created three otherwise comparable experimental conditions with inter-trial interval *SD*s of 88 ms, 176 ms, and 351 ms.

The data for this experiment were collected under similar but not identical circumstances to the previous experiments. The primary difference was using a computer monitor with a slower raster refresh rate of 72Hz. This difference resulted in small changes in the presentation durations of the fixation and response stimuli. Otherwise the apparatus and procedures were kept as similar as possible to the previous studies although the timing of the displays was adjusted to accommodate the 72Hz monitor.

Each simple response time trial began with a ready signal (+++) displayed in the center of a computer monitor. The ready signal was visible for 181 ms and was replaced after this fixed time interval by a blank screen. After 700 ms, the blank screen was replaced by a signal to respond (#####) presented in the center of the computer monitor. The signal to respond remained visible until a response was recorded.

The duration of the blank screen between the offset of the signal to respond and the onset of the ready signal was manipulated, i.e., the ITI. The mean ITI was always 700 ms, but the amplitude of external variation was manipulated by increasing the standard deviation of the inter-trial intervals across the experimental conditions. Participants responded by clicking a joystick trigger-button. They completed 64 practice trials within approximately 3 min and completed 1100 trials in about 20 min.

Results

Spectral analyses require pre-preparation of data, trimming of extreme values and normalization, as noted already. Presently, the

application of trimming criteria to one participant's series yielded fewer than 1024 observations, the minimum number to equate data series for the analyses. That participant's normalized data series was padded with 84 trailing zeros. (This procedure does not impact the values of scaling exponents derived using spectral methods but could impact the results of fractal SDA and DFA analyses; see captions, Figures 5 through 8.) All the analyses yielded the same results irrespective of this one participant's padded data.

Reaction time data in both the 25% Spectrum and the 100% Spectrum yielded reliable differences in scaling exponents due to noise amplitude. Noise amplitude accounted for 14% of variability in scaling exponents from the 25% Spectra, $r^2 = .14$, $F(1, 28) = 4.46$, $p < .05$, and 30% of variability in the scaling exponents of the 100% Spectra, $r^2 = .30$, $F(1, 28) = 11.72$, $p < .05$. However, the scaling exponents of key-release times were unaffected by the magnitude of noise. Figure 9 depicts the outcome of the present experiment, spliced together with the results of the manual response condition of Experiment 4. It illustrates the nature of the inverse relation between the ITI *SD* and the spectral scaling exponents across a range of more than 2 orders of magnitude in the amplitude of the ITI *SD*.

Planned contrasts examined first the interaction between noise amplitude and scaling exponents from 25% or 100% Spectra, which was not statistically reliable, $F(2, 27) < 1$. This null outcome replicates the pattern observed in every experiment we have reported and again fails to support the layering hypothesis. Altogether, this null outcome has held across noise varying in a

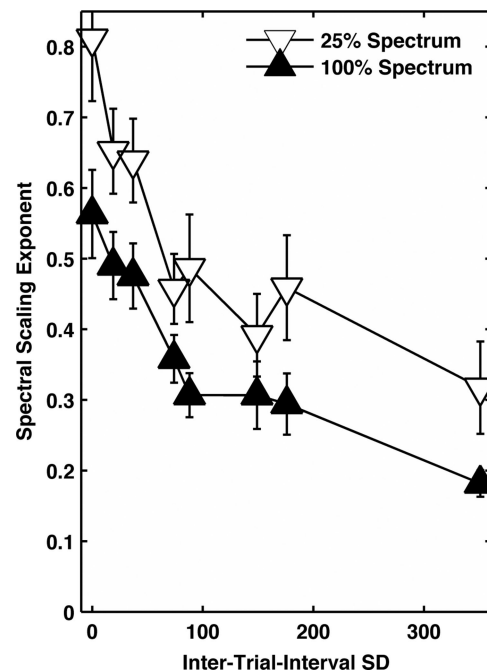


Figure 9. Depicts the Manual results from Experiment 4 spliced together with the results of the wider ITI manipulation of Experiment 5. The X-axis represents the ITI *SD*, the Y-axis track the value of the scaling exponents. The average 25% and 100% scaling exponents derived from the two experiments follow a similar trajectory of progressive whitening, but tend to maintain relatively equivalent differences across the range of manipulated ITI *SD*. ITI *SD* = inter-trial interval standard deviation.

range of over two orders-of-magnitude, noise amplitudes ranging in *SD* from 0 ms to 351 ms.

Of key interest was whether key pressing is dissociated from key releasing behavior, which was examined in the second planned contrast. The implied interaction effect between noise amplitude and response type was indeed statistically reliable, $F(2, 27) = 4.16$, $p < .05$, $\eta^2 = .13$. Scaling exponents of key press times became whiter as noise increased in amplitude, but scaling exponents of key release times did not change reliably despite increases in noise amplitude.

Discussion

Scaling exponents of key release times would reasonably be affected in the same manner as key press times, were the effects of injected white noise due to superficial aspects of measurement procedures. But the observed dissociation of scaling exponents speaks otherwise. That is, to the extent that functional dissociations reliably discover the functional bases of behavior, the effect of injected white noise originates in the functional basis of simple reaction time performance.

The dissociation is conventional evidence that methods of measuring are functionally entangled with the quality of variation in simple reaction times, with observable behavior itself. Yet the dissociation also suggests a way around this conundrum—at least this kind of dissociation has been discussed previously within an alternative workable framework. In the alternative framework, all cognitive and motor performances and other intentional activities emerge as soft assembled devices. Thus, the dissociation of key pressing behavior from key releasing behavior simply distinguishes two temporary and task-specific soft assembled devices, the key pressing device and the key releasing device (Van Orden et al., in press; Kloos & Van Orden, 2010).

Temporary soft assembly supplies devices suited to the temporary requirements of participation in a task environment, reflecting the sometimes fleeting and idiosyncratic configurations of cognitive and motor constraints (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Kugler & Turvey, 1987; Turvey, 1990; Turvey & Carello, 1988). Soft assembly may originate in interaction dominant dynamics, dynamics in which multiplicative feedback among a system's components self-organize behavior (Turvey, 2007; Holden, Van Orden, & Turvey, 2009; Kello & Van Orden, 2009; Kloos & Van Orden, 2009; 2010; Van Orden & Holden, 2002; Van Orden et al., 2003; 2009). In soft assembly, a web of temporary constraints limits the movement possibilities of musculoskeletal tensegrity by forming synergies across the body, which allow the assembled participant to respond as an integrated device, as instructed in the experiment, and to act as an integrated being (Turvey, 2007).

Injected noise perturbs the assembly of key pressing responses unsystematically, resulting in whiter scaling exponents for key press times. Yet each key press creates a stable basis on which to release the key in the details of task demands. Key releasing behavior always follows an organized key press response that damps the injected noise at a contact point. Soft assembly explicitly predicts such detailed dissociations of temporary devices because performance is the coming into existence of behaviors, entrained to specific details of task demands. Even task demands that differ trivially, as in the direction of the finger's motion or by

perturbations of a participant's rhythmic responding, will entrain distinct specialized and dissociable accommodations.

Whitening Injected After Response and Before Ready Signal for Temporal Estimation

The final experiment concerns whether the previous results are peculiar to simple reaction time experiments. Participants in a simple reaction time experiment may be dependent on the timing of trial events to respond as quickly as possible, simply because the functional demands of the task are to respond as quickly as possible. In contrast, temporal estimations are not reactions to a stimulus but deliberate productions, depending instead upon a cognitive judgment about time passing.

The final experiment conceptually replicated and extended the previous methods to a temporal estimation task. Temporal estimation typically yields clear fractal patterns with scaling exponents close to $\alpha = 1$. Also, the key release times were collected, in addition to temporal estimation key press times, to test again for a functional dissociation of key pressing from key releasing. The participant's goal in the temporal estimation task is to respond on each trial by pressing a key, once one second has elapsed, but only after a stimulus appeared. The contingent response on the stimulus signal created an ITI downtime between trials in which to inject noise, similar to the previous reaction time experiments.

Method

Participants. Twenty additional California State University introductory psychology students participated in exchange for course credit. Ten undergraduate students were assigned randomly to each of two conditions.

Apparatus and procedure. The apparatus was identical to those of the previous studies. We changed the details of the display and instructions to be consistent with a temporal estimation protocol. The fixation stimulus and the ISI were eliminated from the display and replaced with equivalent intervals of a blank-screen display (601ms total, 181ms + 420ms, respectively). Two ITI conditions were used. The first used a constant 700-ms ITI. The second implemented the variable ITI condition that yielded an average ITI of 700-ms with a *SD* of 351ms, as described in Experiment 4. Participants were asked to press a joystick button once they believed the stimulus (i.e., #####) had been displayed on the screen for 1 s.

Results

The data series was prepared as in the previous experiments using different trimming criteria, which eliminated temporal estimates and key release times that were faster than 10 ms or slower than 5 s, accommodating the much longer response times (compared to simple reaction times). One participant held the joystick in an idiosyncratic position such that 23% of the key release times were recorded as zeros.

The same participant produced temporal estimation times like those of other participants. As a precaution, we conducted extra statistical analyses on key release times, including and excluding this participant's spectral exponents, yielding no differences in the pattern of results. With this reassurance, we report summary sta-

tistics that include the idiosyncratic participant's data. Figure 10 depicts the temporal estimation scaling exponents and Figure 11 displays the key-release scaling exponents.

The noise amplitude reliably predicted changes in scaling exponents of temporal estimation for both 25% Spectra, $r^2 = .24$, $F(1, 18) = 5.57$, $p < .05$, and 100% Spectra, $r^2 = .25$, $F(1, 18) = 5.85$, $p < .05$. Noise amplitude did not reliably correlate with scaling exponents of key release times, however, $r^2 = .06$, $F(1, 18) = 1.07$, NS, for 25% Spectrum; $r^2 = .09$, $F(1, 18) = 1.87$, NS, for 100% Spectra. These planned comparisons led us to expect a statistically reliable interaction effect between noise amplitude and the scaling exponents from the two measurements.

Nonetheless, a mixed 2 (ITI SD) \times 2 (Measure Type) ANOVA, to test for the interaction, came up dry; the interaction was not statistically reliable. The scaling exponents of key release times displayed a trend toward whiter scaling exponents, weakly parallel to the reliable trend in the scaling exponents of the temporal estimation data. Thinking in terms of soft assembly, is it possible that the key releases have become functionally relevant within the temporal estimation task?

Yes, it's possible. A temporal estimation task places a premium on responding with regular consistent time estimates, all equal in duration. Perhaps participants were more tuned to the timing of trial events to facilitate this goal. Alternatively, perhaps short ITIs sometimes encroach upon the moment when the key is first released but is not yet fully released, prompting the next trial's temporal estimation before the finger has left the key. If so, then the key release may become enfolded in the timing of the oncoming trial. This perturbation may be most likely when partici-

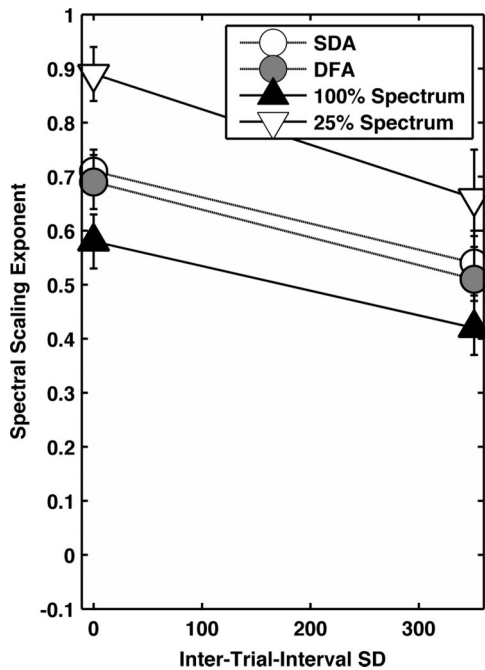


Figure 10. The average scaling exponent values for the temporal estimation series, using the 25% and 100% Spectrum methods, as well as the SDA and DFA methods. The X-axis indicates the value of the ITI SD, the Y-axis tracks the average values of the scaling exponents. ITI SD = inter-trial interval standard deviation.

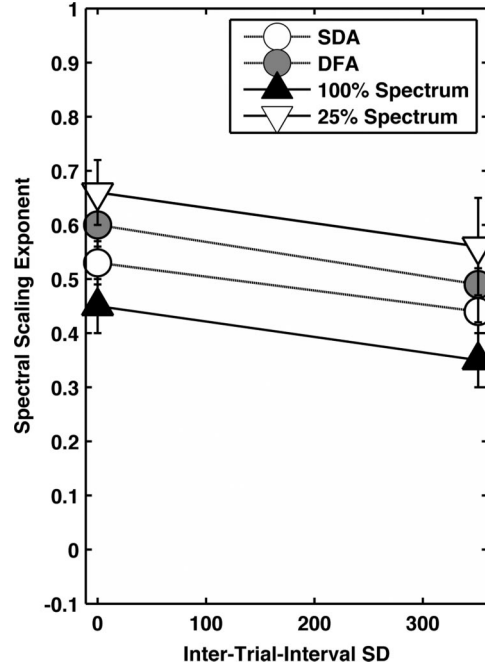


Figure 11. The average scaling exponent values for the key-release times from the temporal estimation series, using the 25% and 100% Spectrum methods, as well as the SDA and DFA methods. The X-axis indicates the value of the ITI SD, the Y-axis tracks the average values of the scaling exponents. ITI SD = inter-trial interval standard deviation.

pants expect mostly very long ITIs, as in large amplitude noise conditions. A separate research project, so far, favors the latter hypothesis.

Discussion

Previously, Wagenmakers et al. (2004) reported a weak fractal pattern in a temporal estimation task in contrast to Gilden et al. (1995) who reported a robust fractal pattern. Gilden and his colleagues had used a self-paced temporal estimation, reducing task demands by giving the participant control over trial progression. The present experiment included a baseline condition that, in effect, substituted a fixed ITI condition for self-pacing and introduced the task demand to harness temporal estimations to the predictable pace of task trials. Wagenmakers et al., in turn, substituted unsystematic ITIs for our fixed ITI condition, introducing a task demand to harness temporal estimation to a randomly determined pace.

If these respective task demands form a rank order of demand (in some sense)—self-paced ITI, predictable fixed duration ITI, unpredictably changing ITIs—they also correctly predict the rank order of outcomes: scaling exponents closest to $\alpha = 1$ for Gilden et al., closest to $\alpha = 0$ for Wagenmakers et al., and our baseline conditions α values in the middle. Altogether, these outcomes suggest that estimates of temporal intervals are not distinct acts by participants; they are neither isolated productions nor punctate events. Temporal estimates instead reflect the experimenter-determined possibilities for coupling by the participant to the timing demands of the task protocol (Balasubramaniam, 2006;

Delignières, Lemoine, & Torre (2004); Kloos & Van Orden, 2010; Schöner, 2002). The measured behaviors are more tightly coupled to task timing and other demands than has been otherwise anticipated.

General Discussion

No single set of experiments can rule out such fundamental distinctions as those between *signal* and *noise*, *competence* and *performance*, or *mind* and *behavior*, and the present experiments should be met skeptically in that light. Another misguided conclusion would be to mistake the present findings as support for a kind of *behaviorism*. The historical frameworks that are sometimes lumped together and called behaviorism also assumed a layering hypothesis about signal versus noise in data. Deeper questions are at issue here: in the relations between what exists (*ontology*), what we can know (*epistemology*), and how that which we can know is distinguished in the “basic experience of the world” (*phenomenology*, Merleau-Ponty, 1962, p. viii).

The conundrum of our title stems from the facts now in evidence: superficial white noise changes the kind of variation in measurements, and the kind of variation in measured values is the basis for how to think about the nature of behavior. The four experiments reported here all manipulated minor details in the circumstances of measurement, and reliably changed the kind of variation in behavioral measurements. Most important, all four experiments failed to distinguish exogenous noise from endogenous signal in contrasts between scaling exponents from 25% and 100% Spectra. All four experiments failed to distinguish superficial sources of variability from fundamental sources.

We noted a weakness of the present logic, stemming from the unknown relation between the magnitudes of empirical noise and the hypothetical impact of noise amplitudes on scaling exponents. It remains possible that an interaction effect of layered noise with behavioral signals yet lies beyond the range of the present manipulations. In that regard, the present manipulations of noise amplitude span over two orders of magnitude, exaggerating standard methods of control that introduce random noise. On this basis, from here on, we will take seriously that we have failed to distinguish endogenous from exogenous variation in behavioral measurements.

In light of that outcome, one experiment also clearly dissociated key pressing behavior from key releasing behavior, conventional evidence for separate and independent cognitive devices, one deciding when to press a key and the other deciding when to release a key. Trivial task devices such as these derive from task idiosyncrasies that vary across laboratory trials. Constraints due to task demands provide a basis for soft assembly of task devices from human participants as we have discussed.

Continuing in this vein, the functional dissociation of scaling exponents, key press times from key release times, and decision devices of key pressing and releasing, demonstrate a rarely considered means to change behavior: Mind and body may functionally reorganize in behavior. In this regard, the flexible changes generally observed of scaling exponents indicate sufficiently flexible intrinsic dynamics to support functional reorganization (Kiefer et al., 2009; Riley, Shockley, & Van Orden, in press; Warren, 2006). Alternatively, if we had supported the additive layering

hypothesis, the evidence would have suggested that cognition and behavior were simply too stable to reorganize on-line.

Models and Statistics

Previously, Thornton and Gilden (2005) successfully modeled changes in scaling exponents, similar to those observed here, as the layered sum of fractal variation and white noise. Their description of variation in performance also inspired the present manipulation of injected white noise. Their method of fitting spectral slopes, using $1/f$ noise plus white noise, would also successfully mimic the present data outcomes. We have not challenged the capacity of their method to fit data. Our concern, instead, is about a larger debate about tools to decide the nature of human performance.

Purely statistical tools are currently used as decision aids to supply objective answers about whether one model is superior to another and specifically to decide whether data truly contain fractal noise (see also Lemoine, Torre, & Delignières, 2006; Wagenmakers et al., 2005; Farrell et al., 2006b). As we see now, the matter cannot be decided in such a starkly mechanical manner (see also Gilden, 2009). The present outcomes illustrate both the limits and promise of decision making with statistical techniques: The decision is only as reliable as the technique’s underlying assumptions, but one can test the assumptions that underlie the technique.

Statistical tools implicitly codify their assumptions in the phenomenon of interest. Thus, while powerful aids, statistical tools are insufficiently flexible to anticipate all the practical uncertainties in scientific discourse (Hand, 2006). These facts are especially salient when a science moves into a new domain of inquiry and a new discourse, and cognitive and behavioral sciences have recently joined the discourse of complexity science and nonlinear dynamics.

The same practical concerns apply to the tools used here. Practically speaking, standard spectral and fractal tools have important strengths and can reveal valuable information, though each tool has its blind spots. Like any statistical technique, spectral and fractal methods are susceptible to artifacts (Caccia et al., 1997; Press, Teukolsky, Vetterling, & Flannery, 1992; Holden, 2005). However, weaknesses in one context may be strengths in another—we capitalized on “weaknesses” of spectral methods to construct a sensitive test of the layering hypothesis for instance.

Traditional fractal analyses are sensitive to changes in the fractal pattern across conditions and participants (see captions, Figures 5 through 8), and changes in scaling exponents provide the more reliable basis for conclusions about performance dynamics—we recommend a focus on how they change instead of their absolute values (Van Orden et al., in press). Yet the present fractal methods assume that data series are monofractal, having the same fractal dimension throughout. They do not adequately characterize more complex multifractal structure, for instance.

By analogy, a quantitative summary statistic such as the sample standard deviation supplies a frozen snapshot of variation. It supposes that a single overall quality of variation is sufficient to summarize the local details of variation across many successive measurements. In other words, it requires that the local timing and the local contexts of a measurement, and the local qualitative changes in a participant, play negligible roles in determining the

outcome of a measurement. Monofractal statistics make the same assumption.

But data like those presented here are not monofractal; they are multifractal; and so we confront meaningful limits of the present statistics (Ihlen & Vereijken, 2010). Multifractal statistics summarize changes in scaling exponents and fractal dimension that accrue during data collection. The central tendency of multifractal statistics is closely approximated by a monofractal scaling exponent, however the range of variation in scaling exponents can be wider, narrower, or not apparent at all, depending on idiosyncrasies of task and participant.

All the present data are reliably multifractal, but they do not differ reliably in the range of variation, only in their central tendency. This outcome allowed the present focused discussion of monofractal scaling exponents. Nevertheless, the findings of Ihlen and Vereijken (2010), that trial-series of response times are generally multifractal, is telling in several respects. On one hand, their findings underscore the need to bring into play all the reliable dynamical signatures available in human performance. On the other hand, their findings represent a different kind of evidence for the conundrum of our title (e.g., see Van Orden, Kloos, & Wallot, in press).

The Conundrum

Taking into account what is presently known about variation in human performance, long traditions of analysis of human performance have been false and misleading; they were framed too rigidly around simplifying assumptions. After all, the “meaningless” structure of inter-stimulus and inter-trial downtimes changes the kind of response behaviors that we observe.

Human performances are not the static objects portrayed in conventional statistics. Human performances are sufficiently fluid and accommodating that they may change in quality to meet subtle idiosyncratic requirements of a task at hand, as well as to meet on-line challenges within the task. Although practical purposes may sometimes justify glossing over these facts, the forthright purposes of basic science cannot.

The accumulated evidence could mean that a tighter coupling exists between the participant and the measurement environment than has been traditionally assumed. It is even plausible that the pattern of variation gauges the capacity of a participant to entrain to the temporal unfolding of task environments; and this pattern of entrainment is what is fluidly reorganized (Kloos & Van Orden, 2010; Van Orden et al., 2009). If so, then cognitive scientists may confront issues like those that motivated quantum mechanics.

One issue discussed elsewhere is *complementarity*, which concerns the different perspectives on behavioral phenomenon yielding different ideas about behavior (Atmanspacher, Römer, & Walach, 2002; Flach, Dekker, & Stappers, 2007; Uttal, 2007). On the one hand each behavioral datum represents a singular punctate event. On the other hand each datum participates in the larger fractal wave of $1/f$ noise. These facts remind us of complementarity and the electron (Van Orden et al., 2010). On the one hand an electron behaves as a singular distinct particle. On the other hand each electron participates in larger patterns of wave interference.

Complementarity forced physicists to rethink the relationship between the act of taking a measurement and the outcome of the measurement that was taken. They concluded eventually that the

particular circumstances of measurement are intimately coupled to phenomena like electron behavior, and play a fundamental role in what is finally observed. That is, when the coupling between research methods and observations is sufficiently tight, then scientists confront *entanglement*, which concerns the wholeness of a tightly coupled system, and which is destroyed in the taking of a measurement (Atmanspacher et al., 2002).

If psychology truly confronts entanglement and complementarity, then such issues have come full circle. Niels Bohr was tutored in complementarity by the psychologist Arthur Rubin and famously illustrated complementarity using “belief” and “doubt,” an example from William James (Atmanspacher, in press). Of course, cognitive and behavioral sciences are not simply quantum mechanics. Psychological phenomena may parallel phenomena in physics but they cannot be equated to quantum phenomena. All that we claim is that these two sciences could both confront entanglement of measurement technique with measured phenomena—the conundrum of our title.

References

- Atmanspacher, H. (in press). Quantenphysik und quantenalltag. In U. Gehring (Ed.), *An den Grenzen des Wissens*. Paderborn: Fink Verlag. (Trans. by the author).
- Atmanspacher, H., Römer, H., & Walach, H. (2002). Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics*, 32, 379–406.
- Balasubramaniam, R. (2006). Trajectory formation in timed rhythmic movements. In M. L. Latash & F. Lestienne (Eds) *Motor control and learning* (pp. 47–54). New York: Springer.
- Bassingthwaight, J. B., Liebovitch, L. S., & West, B. J. (1994). *Fractal physiology*. New York: Oxford University Press.
- Buhrer, M., Sparrer, B., & Weikunat, R. (1987). Interval timing routines for the IBM PC/XT/AT microcomputer family. *Behavior Research, Methods, Instruments, & Computers*, 19, 327–334.
- Caccia, D. C., Percival, D., Cannon, M. J., Raymond, G., & Bassingthwaight, J. B. (1997). Analyzing exact fractal time series: Evaluating dispersive analysis and rescaled range methods. *Physica A*, 246, 609–632.
- Delignières, D., Lemoine, L., & Torre, K. (2004). Time intervals production in tapping and oscillatory motion. *Human Movement Science*, 23, 87–103.
- Eke, A., Hermán, P., Bassingthwaight, J. B., Raymond, G. M., Percival, D. B., Cannon, M., . . . Ikrényi, C. (2000). Physiological time series: Distinguishing fractal noises from motions. *European Journal of Physiology*, 439, 403–415.
- Eke, A., Hermán, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, 23, R1–R38.
- Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2006a). $1/f$ noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic Bulletin & Review*, 13, 737–741.
- Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2006b). *Methods for detecting $1/f$ noise*. Retrieved from <http://eis.bris.ac.uk/%7Efpssaf/tgreplysims.pdf>
- Flach, J. M., Dekker, S., & Stappers, P. J. (2007). Playing twenty questions with nature (the surprise version): reflections on the dynamics of experience. *Theoretical Issues in Ergonomics Science*, 9, 125–154.
- 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, 1441–1467.

- Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). $1/f$ noise in human cognition. *Science*, 267, 1837–1839.
- Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science*, 21, 1–14.
- 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 from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>
- Holden, J. G., Van Orden, G. C., & Turvey, M. T. (2009). Dispersion of response times reveals cognitive dynamics. *Psychological Review*, 116, 318–342.
- 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.
- Järvilehto, T. (1998). The theory of the organism-environment system: I. Description of the theory. *Integrative Physiological and Behavioral Science*, 33, 321–334.
- Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of $1/f$ scaling in speech reflects the metastable basis of cognition. *Cognitive Science*, 32, 1217–1231.
- Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, 136, 551–568.
- Kello, C. T., & Van Orden, G. C. (2009). Soft-assembly of sensorimotor function. *Nonlinear Dynamics, Psychology and Life Sciences*, 13, 57–78.
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 812–832.
- Kiefer, A. W., Riley, M. A., Shockley, K., Villard, S., & Van Orden, G. C. (2009). Walking changes the dynamics of cognitive estimates of time intervals. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1532–1541.
- Kloos, H., & Van Orden, G. (2010). Voluntary control of cognitive and motor tasks. *Mind & Matter*, 8, 19–43.
- Kloos, H., & Van Orden, G. C. (2009). Soft-assembled mechanisms for the unified theory. In J. P. Spencer, M. Thomas, & J. McClelland (Eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered* (pp. 253–267). New York: Oxford University Press.
- Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lemoine, L., Torre, K., & Delignières, D. (2006). Testing for the presence of $1/f$ noise in continuation tapping data. *Canadian Journal of Experimental Psychology*, 60, 247–257.
- Mandelbrot, B. B., & Wallis, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Research*, 5, 321–340.
- Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge.
- Peng, C.-K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in non-stationary heartbeat time series. *Chaos*, 5, 82–87.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Riley, M., Shockley, K., & Van Orden, G. (in press). Learning from the body about the mind. *Topics in Cognitive Science*.
- Schöner, G. (2002). Timing, clock, and dynamical systems. *Brain and Cognition*, 48, 31–51.
- Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin and Review*, 12, 409–441.
- Turvey, M. T. (1990). Coordination. *American Psychologist*, 45, 938–953.
- Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, 26, 657–697.
- Turvey, M. T., & Carello, C. (1988). Exploring a law-based, ecological approach to skilled action. In A. M. Colley & J. R. Beech (Eds.), *Cognition and action in skilled behaviour* (pp. 191–203). Amsterdam: North Holland Press.
- Uttal, W. R. (2007). *The immeasurable mind: The real science of psychology*. Amherst, MA: Prometheus.
- Van Orden, G. C., & Holden, J. G. (2002). Intentional contents and self-control. *Ecological Psychology*, 14, 87–109.
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132, 331–350.
- Van Orden, G. C., Holden, J. G., & Turvey, M. (2005). Human cognition and $1/f$ scaling. *Journal of Experimental Psychology: General*, 134, 117–123.
- Van Orden, G. C., 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. C., Kloos, H., & Wallot, S. (in press). Living in the pink: Intentionality, wellbeing, and complexity. In C. Hooker (Ed.) *Philosophy of complex systems. Handbook of the philosophy of science*, (Vol. 10.). Amsterdam: Elsevier.
- Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, 25, 111–172.
- Voss, R. F. (1992). Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences. *Physical Review Letters*, 68, 3805–3808.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f$ alpha noise in human cognition. *Psychonomic Bulletin & Review*, 11, 579–615.
- Wagenmakers, E. J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General*, 135, 108–116.
- Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.
- Warren, W. (2006). The dynamics of perception and action. *Psychological Review*, 113, 358–389.
- Woodworth, A. S., & Schlosberg, H. (1954). *Experimental psychology, revised edition*. New York: Holt, Rinehart, and Winston.

Received December 4, 2009

Revision received June 6, 2010

Accepted July 10, 2010 ■