CHAPTER 6

Gauging the Fractal Dimension of Response Times from Cognitive Tasks

John G. Holden

Department of Psychology
California State University, Northridge
Northridge, CA 91330-8255
U. S. A.
E-mail: jay.holden@csun.edu

An unexpected and exotic brand of variability resides in the trial-by-trial fluctuations of human judgments of passing time. The pattern, called $1/f$ or pink noise, is a construct from fractal geometry. Pink noise is associated with complex systems whose components interact on multiple time scales to self-organize their behavior (Bassingthwaighte, Liebovitch, & West, 1994; Jensen, 1998; Van Orden, Holden, & Turvey, 2003; see also Aks, Chapter 7). This chapter describes the phenomenon of pink noise, and explains how to conduct statistical analyses that identify it in response time data from elementary cognitive tasks.

It was Gilden, Thorton, and Mallon (1995) who first reported pink noise in response time variability during the fundamentally cognitive task of estimating fixed intervals of time. Gilden et al.'s temporal estimation task required participants to repeatedly estimate fixed intervals of time—in essence to "become a clock"—by pressing a button at each instant they believed a specific time interval had elapsed. Separate laboratory sessions were administered for each of several fixed target time-interval conditions. In each session 1000 time-interval judgments were collected in succession. The time-interval conditions ranged from 1/3 s up to 10 s. Of course, no participant's succession of time-interval judgments was exactly the same on every trial. Instead they varied from trial to trial. Lining up the series of successive time estimates in the strict order in which they were collected (i.e., trial 1, 2, ... 1000) yielded a *trial series* of response times, which was treated very much like a standard *time series* in Gilden et al.'s statistical analyses. Pink noise was revealed in the intrinsic residual variability that remained after the average time

interval that each participant produced, for each target interval, was removed from each trial series. Thus, pink noise emerged in the structure of the "background noise" of cognitive performance—the intrinsic variability of a person's judgments of passing time.

To begin to understand the phenomenon of pink noise it is perhaps easiest to simply examine it visually. Figure 6.1A displays an example of an individual participant's trial series from a temporal estimation task that used a method similar to that of Gilden et al. (1995). The x-axis depicts the successive trials in the experiment and the y-axis records the participant's time estimate on each trial, in terms of standard deviations from their average time interval, taken across all trials (i.e., the time estimates are normalized and graphed as z-scores).

The overall pattern of trial-to-trial fluctuation is consistent with pink noise. Notice the undulating "waves" of relatively longer and then shorter time estimates that travel across the series. One shorthand way to describe the overall rising and falling trends is to say that the first 500 or so trials follow a giant inverted-U shape or an arc. It is fair to say that a similar large arc begins around the 500th trial and continues to the end of the series of observations. Now, look within each large arc, and similarly shaped arcs which run across fewer trials, perhaps only 50 to 100 trials at a time, can be discerned. Inside the smaller arcs are even smaller ones, and so on. Loosely speaking, the trial series is comprised of a progression of nested, similarly shaped arcs or patterns of fluctuation. Of course, there is nothing special about the arcing inverted-U shape; you could imagine M or W shapes, or even right-side up Us, for example. What is important in this example is the concept of a shape or pattern that is comprised of smaller copies of essentially the

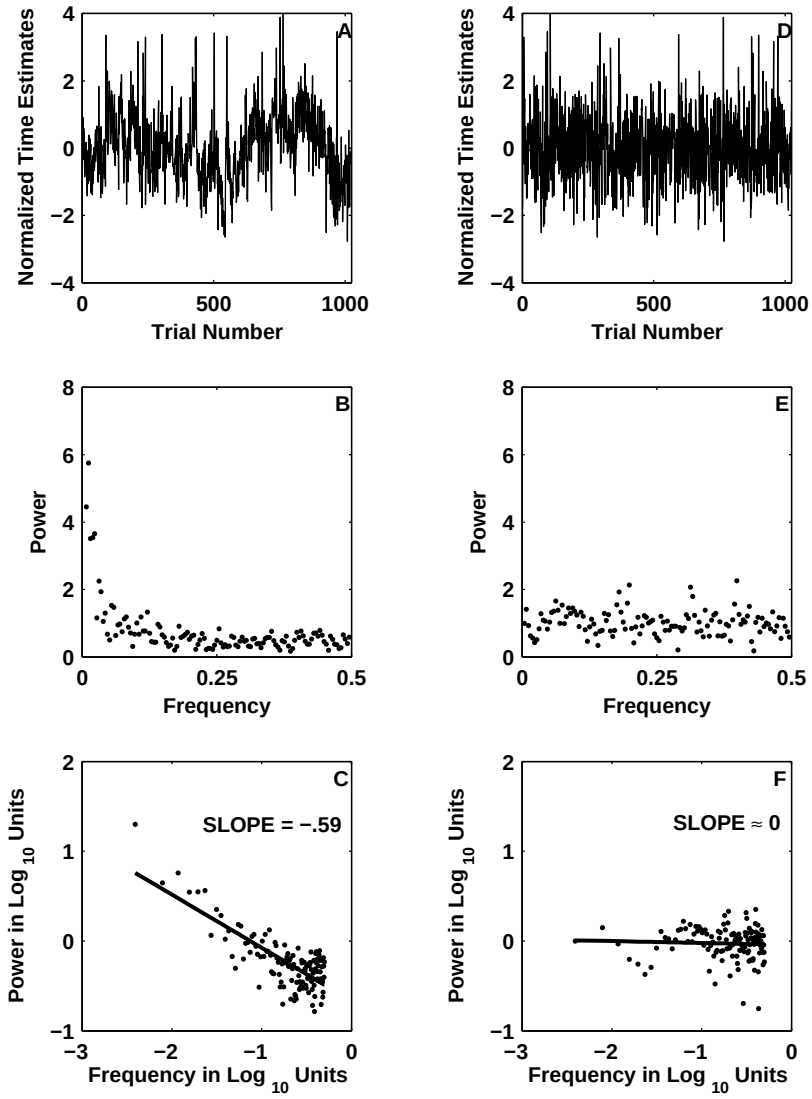

Figure 6.1. (A) A trimmed, detrended and normalized trial-series of 690 ms time interval estimates for a single participant. The x-axis indexes the successive trials in the experiment, the y-axis indexes the time interval judgments, relative to the overall mean and standard deviation of the trial series. (B) Results of a power spectral analysis, in linear units. Frequency is plotted on the x-axis, from lowest (near the origin) to highest. The y-axis indexes the power or relative energy of each frequency. Larger values indicate more power. (C) Results of the same spectral analysis plotted in (B), on double-logarithmic axes. The negatively accelerated linear relation, with a slope less than 0 but greater than –1, is consistent with pink noise. (D)-(F) The same succession of plots as (A) through (C) for a randomly reshuffled "surrogate" version of the data depicted in (A). Random reshuffling destroys the natural trial ordering and yields white noise. (E) depicts the spectral analysis of the white noise on a linear scale. There is no systematic variation in power as a function of frequency. This power spectrum is characteristic of white noise; all frequencies have roughly equal power. (F) depicts the same power spectrum on double-logarithmic coordinates. The slope of the regression line is approximately zero, which is consistent with white noise.

same shape—the notion of a nested structure of similar-shaped fluctuations.

If you are not sure you see this structure, simply compare Figure 6.1A with Figure 6.1D, which represents exactly the same data set depicted in Figure 6.1A, but where the order of the successive data points was randomly shuffled. The shuffling procedure destroyed the nested, statistically self-similar pattern of trial-to-trial fluctuations characteristic of pink noise. The random rearrangement of the series yields a pattern called *white noise*. Notice that just about any portion of the shuffled data series depicted in Figure 6.1D could be used as relatively good "stand-in" for any other portion of the series. This is not true for the trial-ordered data plotted in Figure 6.1A, for which most of the observations between trials 500-600 fall below the overall mean, while the majority of the observations between trials 800-900 fall above the overall mean.

The two different arrangements of the same data set are quite distinct, illustrating the difference between pink noise and white noise. For pink noise, the local means and standard deviations depend on where in the series the sample was taken. White noise indicates statistical independence from observation to observation, and local sample means and standard deviations do a good job of describing other local samples, and an overall population mean. This fact about white noise forms the cornerstone of inferential statistics, such as t-tests, analysis of variance (ANOVA), and regression.

In the context of response time research *pink noise* refers to a statistically self-similar (see Liebovitch, Chapter 5) pattern of trial-to-

trial variability. The pattern is structured such that persistent, long-term fluctuations across several hundred trials nest within themselves progressively smaller, proportionately scaled fluctuations across a few decades of trials. Nested within those fluctuations, one finds even smaller patterns of fluctuation, and so on. Pink noise is unexpected from the perspective of conventional statistical intuitions inherited from the standard linear statistical tools of behavioral science research. Those intuitions lead to an expectation that successive individual time judgments, decisions, or other elementary cognitive performances should vary unsystematically or randomly from trial to trial. That is, individual observations are assumed to be statistically independent. After all, it is properties of the presented stimulus that are conventionally assumed to be driving a participant's response, not aspects of the previous response.

Since Gilden et al.'s (1995) report, pink noise was uncovered in the trial-by-trial variability, or trial-series, of a wide range of standard cognitive psychology laboratory tasks that collect response times. Examples include simple reaction time, speeded word naming, choice reaction time, lexical decision, and mental rotation, among many others (e.g., Gilden, 1997; Kelly, Heathcote, Heath, & Longstaff, 2001; Van Orden et al., 2003). The pattern is not limited to cognitive activities; pink noise appears in measurements of human performance that use dependent measures other than response time, such as patterns of eye movements (Aks, Zelinsky, & Sprott, 2002), postural sway (Riley, Wong, Mitra, & Turvey, 1997), and self-reports of changes in mood over time (Delignières, Fortes, & Ninot, in press; see Gilden, 2001, and Van Orden et al., 2003, for reviews).

The goal of this chapter is to provide a primer to the geometric concepts and statistical techniques that are necessary to characterize pink noise in trial series of response time measurements derived from cognitive performances. The first step is to introduce three interrelated concepts from fractal geometry: *Self-similarity*, *scaling*, and *fractal dimension*. Those ideas motivate statistical analyses that are aimed at the identification of fractal patterns in data from empirical phenomena. A description of a simple temporal estimation task, modeled in large part after the method used by Gilden et al. (1995), follows the introductory sections. A trial-series resulting from the illustrative temporal estimation task is used to provide a practical context for a tutorial presentation of the statistical procedures involved in a fractal analysis of response time data, including spectral density estimation and fractal dimension estimation. Potential theoretical implications of pink noise are briefly discussed in the final section of the chapter.

FRACTAL PATTERNS

Self-Similarity and Scaling

Two key constructs in fractal geometry are pattern and self-similarity of pattern. The parts of fractal objects are composed, in some way, of copies of the whole object (Feder, 1988; Mandelbrot, 1982; see Liebovitch, Chapter 5). Ideal geometric fractals may be composed of exact replicas of the whole object—they are strictly self-similar. By contrast, statistical fractals are *self-affine*, or *statistically self-similar*; they are composed of statistically equivalent replicas of the whole object. Naturally occurring fractals usually exhibit statistical self-similarity.

A contrast between the left and right sides of Figure 6.2 illustrates this distinction. The left side of the figure depicts a Sieripinski Gasket, a classic mathematical fractal. It was generated by removing a white, smaller triangle, with vertices that fall at the midpoints of the sides of the largest, gray outer "initiator" triangle. Next, smaller similar white triangles were removed from the three new triangles that were formed, and their centers were, in turn, removed. The construction process can be continued indefinitely (see Peitgen, Jürgens, & Saupe, 1992, for details). The Sieripinski Gasket is comprised of smaller, exact copies of itself—it is strictly self-similar. The top panel on the right side of Figure 6.2 depicts 8192 observations of idealized pink noise, a statistical fractal. The middle panel "zooms in" on the center 4096 observations of the same series depicted in the top panel, and the bottom panel depicts the center 2048 observations that appeared in both the middle and top panels. When the x- and y-axes are appropriately scaled, the pieces of the series are visually and statistically indistinguishable from the overall series—pink noise is statistically self-similar, or *self-affine*.

Fractal patterns in nature are composed of nested forms that cannot be measured on a single scale of measurement. The result of a measurement depends on the scale, or *size* of the increment used to take the measurement (Bassingthwaighte et al., 1994; Mandelbrot, 1982; Schroeder, 1991; see also Liebovitch, Chapter 5). For example, the measured length of the British coastline increases proportionally as the scale of the "yardstick" used for measurement is shortened from kilometers to meters. An even shorter, centimeter scale of measurement would result in a further proportional increase in the

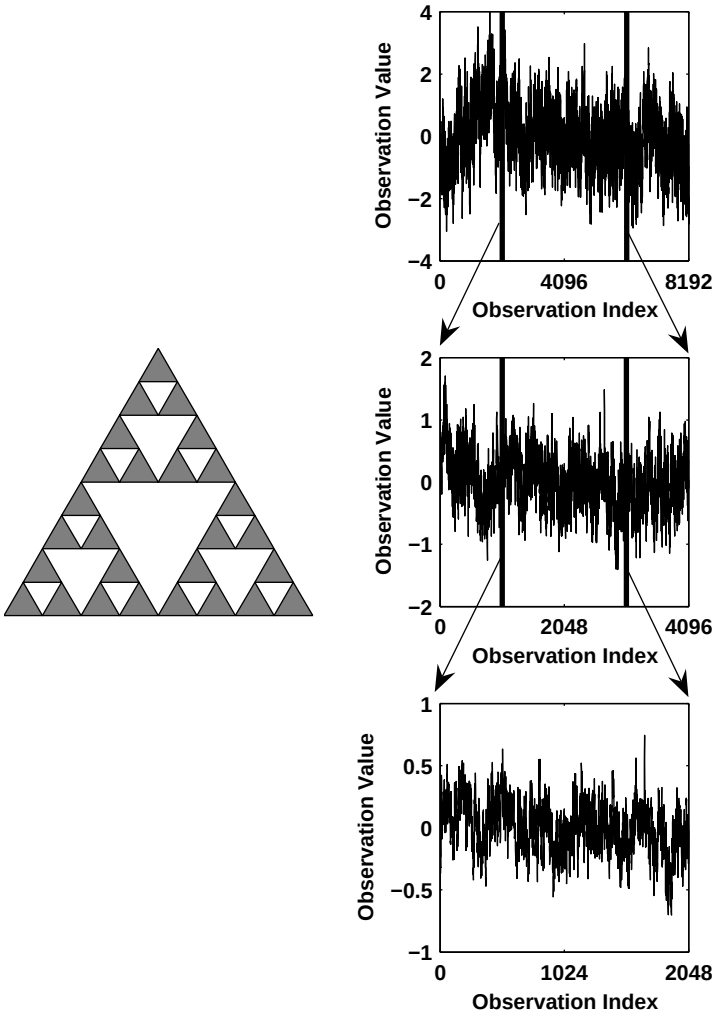

Figure 6.2. The object depicted on the left side of the figure is a classic self-similar mathematical fractal called a Sieripinski Gasket. It is generated by removing successive generations of triangles (white) from the centers of the gray triangles. The three panels on the right side of the figure illustrate how idealized pink noise is *statistically* self-similar. The middle panel zooms in on a piece of the series depicted in the top panel, and the lower panel zooms in on a piece of the series depicted in the middle panel. Each magnification of the pieces of the series results in a new series that looks essentially the same as the original series.

measured length of the coastline. The changing measurements arise as a consequence of using regular line segments, the yardstick, to approximate the irregular, nested, self-similar structure of coastal bays

and peninsulas. The length measurement increases when a bay or peninsula that was not captured at a lower resolution adds length at a higher resolution. Thus, as smaller and smaller sub-bays and sub-peninsulas are resolved, they add to the length of the coastline.

When measurements change as a function of measurement scale, there is no "true" or *characteristic* value for the measurement. The length of the British coastline grows in proportion to the precision of the yardstick used to measure length. This proportional power-law scaling relation between the size of the yardstick and the costal length implies that results of a measurement procedure depend on the measurement scale or sampling unit used to take the measurement (over a finite range of scales). A *power-law scaling relation*, a linear relation between the logarithm of the scale and the logarithm of the measurement result, is commonly observed in natural fractal phenomena, and is symptomatic of self-similar patterns (Bassingthwaighte et al., 1994, Peitgen, Jürgens, & Saupe, 1992). It is the functional form of the scaling relation that, in turn, is used to describe and even model subtly different aspects of coastlines (e.g., Mandelbrot, 1982).

Scaling in Statistics

What is measured with a statistical sample depends on what statistic is computed on the sample of observations. A sample mean is a measure of location—the center or balance point of a distribution of observations. If the variability in the sample is unsystematically but symmetrically dispersed about the mean, and the observations are statistically independent of one another, then a sample mean identifies

a location on the number line of the dependent measure that best characterizes the level or amount of the measurement in the context of sampling error. For instance, mean response time is used to estimate the duration of time that passes between the presentation of a stimulus and the collection of a response in a response time task. However, response time distributions are typically positively skewed, and the pattern of pink noise implies that successive observations are not statistically independent.

Thus, patterns of variability constrain the utility of a sample mean as a point estimate of a location or amount. Dispersion measurements, like the standard deviation, are intimately linked to location measurements. They impart information about disagreement among the individual measurements, and indicate how much a measure of location, such as the mean, can be trusted.

Scale magnification makes intuitive sense for an object such as the coastline of Great Britain. The statistical counterpart to a measurement scale is sample size. Obtaining fewer statistical samples corresponds to using measurements at a lower resolution; larger or more numerous statistical samples correspond to measurements at a higher resolution.

At first, the relation between sample size and scale may not be obvious, but remember the example of Great Britain's coastline. The example helps to develop an intuitive analogy about what it means that no characteristic amount of variability may exist in response time data. The coastline scaling relation indicates that smaller spatial features are nested within larger spatial features. The key point to keep in mind

about the coastline is the measurement result (the length) depends on the scale used to take the measurement.

Statistically self-similar patterns of fluctuation affect statistical estimates of location and dispersion, like the mean and standard deviation or variance (see also Bassingthwaighte et al., 1994, pp. 33-41). The answer to the question "how long" for a coastline is supplied in terms of a length measurement. The answer to the same question for a response time task is supplied by a parameter estimate, such as mean response time. In contrast to length scales, where the smallest units yield the most accurate measurements, larger sample-sizes—a larger *N*—corresponds to a more precise measurement scale in statistics. This may seem counterintuitive, but it is conventional to assume that certainty in estimates of a population parameter—parameter *resolution*—increases as sample sizes increase (i.e., the central limit theorem).

As mentioned, the ability of a statistic such as the mean to act as a gauge of location depends crucially on the inherent patterns of variability in the data. If the measurements emerge from a process with outputs that conform to standard statistical assumptions, those of the central limit theorem, for example, a sample mean can be trusted to reliably penetrate the variability and reveal increasingly reliable estimates of characteristic values—population parameters—as sample size is increased.

Nevertheless, another possibility exists. Suppose trial-by-trial response time measurements are comprised of a statistically-self similar pattern of positive correlation, where local patterns of

correlation, across just a few successive trials, are nested within increasingly more global and proportionally scaled (enlarged) patterns of positively correlated fluctuations, as broader and broader runs of consecutive trials are spanned. This is essentially a description of pink noise, and just as for the coastline, a change to a more detailed or inclusive scale yields essentially the same pattern that was observed at the less inclusive, lower-resolution scale.

Keeping in mind the analogy to the coastline, it is easier to begin to understand how nested patterns of fluctuation lead to counterintuitive statistical properties. The value of any given lower-resolution *locally* computed mean or standard deviation, that includes just a few adjacent samples, depends crucially on *where* in the trial-by-trial series it was taken—did it come from a waning "bay" or a waxing "peninsula," for instance. Increasing the resolution of the sample by including more and more adjacent observations results in the inclusion of more and larger bays and peninsulas. The implication of each increase in sample resolution is the existence and inclusion of even larger scale fluctuations that reach well beyond the scope of each new, larger scale of resolution. Larger samples simply admit a broader range of variability. The resulting increase in variability created by widening the window of observation may *outpace* a sample statistic's ability to stabilize about a particular *characteristic* value, in the normal way, through the process of aggregating larger and larger samples.

As such, the utility of the mean and standard deviation as simple measures of location and dispersion may be foiled because of the proportional, nested patterns of fluctuation. Different sample means and standard deviations, taken at different locations or times, would

tend to disagree in their location and dispersion estimates (up to the limits of the system or sample), resulting in a persistent heterogeneity, or disagreement among the sample statistics themselves. In this way, patterns of variability that are comprised of nested, interdependent, statistically self-similar fluctuations hamper the ability of these statistics to provide a uniform summary or gauge of certain kinds of data sets. Simply put, the concept of a *summary* statistic is not neatly applicable to a heterogeneous process.

All this is *not* to say that descriptive statistics such as the mean and standard deviation are not useful or applicable to fractal or nonlinear science—in many cases they are essential analytical tools. It is really only the semantics of those descriptive measures that is altered in the analysis of a fractal process. Additional and crucial statistical information resides at a higher level, in the manner in which the values of descriptive statistics change as sample size (i.e., measurement resolution) is systematically changed, rather than at the level of any particular summary value that uses a particular sample size. That is, the manner of disagreement across different sample sizes becomes a primary statistical gauge for a fractal analysis.

In the context of response time research, a *fractal dimension* analysis describes how response-time variability scales with sample size. Essentially, it is a statistical analysis that is analogous to taking measurements of a coastline's length using different ruler sizes and reporting how the length changes as the ruler size changes—the fractal dimension analysis determines a scaling relation between sample size and sample variability. The goal of the fractal dimension analysis is to describe the changes in the variability of a measurement across a

range of sample sizes (measurement resolutions) in terms of a power-law scaling relation.

Fractal Dimension

Self-similarity and self-affinity across multiple scales of resolution often gives rise to objects or patterns that occupy noninteger or *fractal* dimensions. Essentially, a *fractal dimension* refers to the spatial dimension of an object whose dimension falls between the standard Euclidean integer dimensions of one, two or three (see Bassingthwaighte et al., 1994). Mandelbrot (1982) explained how the dimension of an object is partly determined by the perspective of the observer (the entry level of the analysis). A tautly stretched piece of thread closely resembles a *line,* an ideal one-dimensional Euclidean object. Tightly weaving the thread, back and forth, results in a piece of fabric, an ideal two-dimensional object. Thus, a *line* can be rearranged so that it occupies *area*. Rolling the thread onto a spool yields an object that occupies *volume* in *3-D space*. Of course, if the spool of thread is viewed from a great distance, its dimension appears to collapse to zero, a *point*.

One way to understand the link between Euclidean geometry and fractal geometry conceptually is to think of fractal geometry as a generalization or elaboration of the standard Euclidean geometry of lines, squares, cubes, and so on. Euclidean objects only occupy integer dimensions, 1 for a line, 2 for a square, and so on. Fractals, however, may occupy *noninteger* dimensions, dimensions that fall in between 1 and 2 or between 2 and 3. How can this be? Refer again to Figure 6.1A, the normalized series of temporal estimates graphed in the

order in which they were collected. They are points connected by a line. Clearly, if every successive time estimate was identical, connecting the points would form a line, and the series would have a Euclidean dimension of one. But any departure from the ideal form of a line begins to occupy or "leak into" the next higher (second, in this case) Euclidean dimension. It is in this spirit that fluctuations in trial-by-trial response times may be said to partly occupy or leak into the next higher Euclidean dimension. In a sense, the variability of time estimates results in the series occupying area, and it will have a dimension between an ideal one-dimensional line and an ideal two-dimensional area. The more jagged and irregular the graph of response times, the more area it occupies. It turns out that this intuitive continuum of relative jaggedness can be characterized formally with the help of a statistical procedure called dispersion analysis. Dispersion analysis results in an estimate of the *fractal dimension* of the trial series. The fractal dimension characterizes the structure of the intrinsic variability in the trial series.

Conventional statistical analyses presuppose that intrinsic variability is *white noise*. White noise yields a jagged and irregular line with a fractal dimension of 1.5, because successive observations are statistically independent of each other. Its fractal dimension indicates the extent to which white noise occupies 2-D space. White noise is uncorrelated noise. By contrast, the successive observations of pink noise tend to be positively correlated. This results in a less jagged trial series, and lower fractal dimensions that fall in the interval between 1 and 1.5.

A Simple Timing Experiment

Data from a timing experiment is now used to demonstrate how to conduct statistical analyses that may uncover fractal patterns in trial series of response times. Except for the setting of trimming criteria, the required statistical techniques are the same for a host of standard cognitive tasks that record response time as a dependent measure. The laboratory protocol was modeled after Gilden et al.'s (1995) temporal judgment task. At the beginning of the experimental session each participant was presented with about one minute's worth of examples of a particular time interval. The time interval was illustrated by repeatedly flashing a simple visual stimulus on a standard PC monitor for the specified period of time. Each participant was then asked to attempt to replicate the example time interval as best as they could, 1100 times, in succession. Participants never received feedback about how accurate they were in their time estimates.

The method section that follows describes a study that replicated Gilden et al.'s (1995) essential finding of pink noise in trial series of temporal estimates. Since the purpose of this chapter is not to disseminate empirical results, but to supply a detailed "how to" tutorial on methods of fractal analysis, just a single participant's trial series is described in the results section (the series that appears in Figure 6.1A was used). The particular series was explicitly selected because it is a very clean and clear example of empirical pink noise; it lacks artifacts that sometimes appear in real-world data and that threaten to further complicate this introductory discussion.

Method

Participants. One undergraduate psychology student participated in exchange for course credit.

Procedure. The participant was given a one-minute sample of a target temporal interval. The sample intervals were constructed by presenting a visual stimulus (########) at the center of a standard CRT monitor controlled by a PC running DMASTR software (Forster & Forster, 1996). The target interval duration was 690 ms or 50 monitor raster refresh cycles. The samples of the 690 ms time interval were generated by displaying the visual stimulus on the monitor for exactly 690 ms, then the monitor went blank for 690 ms, at which point the visual signal again appeared for 690 ms, and so on. The visual stimulus flashed on and off for about one minute. (There is nothing special about the 690 ms duration; the monitor's vertical raster-refresh rate was 72 Hz, or once every 13.8 ms, and 13.8 times 50 equals 690).

On each temporal estimation trial, the visual stimulus (########) was displayed until the participant responded by saying "/ta/" into a microphone, or for a maximum of 10 s. The experimental task was paced by a computer. Each response was followed by a 690 ms inter-trial interval in which the computer monitor was blank. The participant was told to pace her responses so that the visual signal was displayed for the same time interval she saw during the one minute sample time interval session. The participant completed 25 practice trials immediately prior to completing 1100 experimental temporal estimation trials. The entire task took about 30 minutes to complete.

Results

Standard statistical analyses such as regression and ANOVA are typically used in a manner that either ignores the temporal order of the trials in an experiment, or that treats order as a nuisance factor. By contrast, the patterns of fluctuation that unfold across the successive experimental trials are the main focus of the fractal techniques introduced here. Thus, the analysis begins with the participant's trial series of temporal interval estimates *arranged in the order in which they were collected* (Trial 1, Trial 2 …Trial 1099, Trial 1100).

As a practical matter, it is best to present enough trials in an experiment to be left with at least 1024 observations after any timed-out trials, extreme times, and outliers are removed. Presenting 1100 time-estimation trials left a healthy 76 trial "buffer." While it is possible, and potentially informative, to apply fractal techniques to data sets shorter than 1024 observations, the results of the analysis become less reliable as fewer and fewer data points are used. For example, the spectral slopes and the fractal dimension estimates, explained shortly, tend to become more variable as progressively shorter data sets are used (Cannon et al., 1997; Eke, Hermán, Kocsis, & Kozak, 2002). Additionally, the measurements should be collected as regularly in time as possible. A "lined up" series of measurements that were actually collected across different experimental sessions distorts the time scale, and the fractal analysis may not accurately characterize the temporal structure of the series.

Trimming and Detrending. Data trimming procedures are often required to bring the series of time estimates (a finite, irregular natural

object) more in line with the assumptions of spectral and dispersion analyses. The mathematics of spectral analysis assume an ideal, stationary, strictly periodic process of infinite duration. Dispersion techniques are less assumptive and more robust than spectral techniques but nevertheless ultimately assume that the measured process is at least *weakly stationary*—that its mean and standard deviation remain essentially the same over time (Caccia et al., 1997; Chatfield, 1996). If the trial series in hand happens to be a good example of pink noise, then its mean and standard deviation probably do fluctuate as a function of time, or trial.

In general, response time distributions are notorious for the fact that they often contain extreme observations, and the trial series of temporal estimates are essentially response time trial series. No matter their origin, a few extreme measurements or simple long-term trends will likely distort the outcome of a fractal analysis. It is important to note that the main issue surrounding the decision to remove extreme valued data points is not so much whether or not they represent legitimate measurements, as they certainly may. The issue is whether their inclusion will dominate and thus distort the outcome of the analysis.

Response time trial series typically require two censorship passes. The first pass eliminates times that exceed fixed extreme truncation values; different cognitive tasks require the use of different fixed truncation values. Reasonable truncation points can be identified by consulting the relevant literature for typical censorship values. Adopt conservative truncation values from that range as a starting point (err on the side of including more data). When truncating an

observation, just delete it and "close up" the series so that the deleted observation's two immediate neighbors themselves become neighbors. While this procedure slightly disrupts the time-ordering of the series, its overall impact on the analysis is usually minimal. To best preserve the trial order, response times on trials that produced an error in, for example, a choice response time task, should be included in the trial series (see Gilden, 1997).

The main purpose of fixed-value censorship is to facilitate a second pass through the data, which uses the series mean and standard deviation as a censorship origin. For temporal interval estimates, nine observations were less than 100 ms or greater than 3500 ms and were removed. Then the series mean and standard deviation were computed. On the second censorship pass, 13 observations that fell beyond ±3 standard deviations from the series' mean were eliminated.

The dispersion analysis, as described below, *requires* the number of observations to be an integer power of 2 (e.g., $2^{10} = 1024$). While spectral techniques do not always strictly require a series to be an integer power of 2 in length, the algorithms work faster when the data series is an integer power of 2, and some computer implementations of spectral routines do require the series to be an integer power of 2 in length. The two censorship passes eliminated 22 observations, for a total of 1078 remaining observations. The first 54 observations were then eliminated to yield a series that was 1024 observations in length.

Trial series that display self-similar patterns of fluctuation are expected to display nonstationary drift (i.e., trends) at all scales. It can,

however, be difficult to distinguish simple long term trends, or a very-low frequency periodic oscillation from a nested, fractal pattern of long range fluctuations in empirical data sets (Hausdorff et al., 1996). This difficulty again arises from the fact that real data sets have a finite length. A linear trend at the scale of the whole data set could be either a simple linear trend, or a small piece of a fractal pattern of fluctuation that expresses itself across scales that run far beyond the duration of the particular sample of the process at hand. What appears as a linear trend across 1024 observations could be just that, or it could be part of a proportionately scaled fractal fluctuation that runs across 2000-3000 observations. Without the extra data, it is impossible to tell which option is a better description of reality. Only fluctuations that live on scales somewhat smaller than the full length of the series can be resolved clearly enough (i.e. statistically) to determine whether they are consistent with a nested fractal pattern of fluctuation.

Most importantly, simple long-term trends not only have the potential to bias estimates of spectral slopes and fractal dimension, they may also overwhelm the analyses, and yield spurious spectral slopes and fractal dimension statistics (Caccia et al., 1997; Hausdorff et al., 1996). As such, it is prudent to remove at least linear and quadratic trends before conducting the analysis. As a general rule, if the trial series has fractal structure, progressively more liberal detrending procedures will not result in dramatic changes in the overall fractal dimension estimates (Hausdorff et al., 1996). Nevertheless, detrending does eliminate variability at the larger scales, in the neighborhood of the size of the entire series, and any fractal dimension estimation procedure, like those presented later, must be tuned to accommodate

this fact. Otherwise, the detrending introduces its own bias. For the timing trial series trends up to a quadratic were removed. The basic detrending procedure involves generating a least squares linear and quadratic fit to the series, using the index of observation as the x variable and the temporal estimate as the y variable. One way to do this is to use the method of Powered Vectors, in conjunction with Hierarchical Regression (Keppel & Zedeck, 1989), which can be coded on a spreadsheet or accomplished with standard statistical software.

The final preparatory step is to normalize the series to have a mean of zero and a variance of one. That is, transform the data points into z-scores by subtracting the series mean from each observation, and dividing each observation by the series standard deviation (SD). Use the *population* formula for the SD and divide by N, the number of data points in the series, rather than the usual bias-corrected $N - 1$. The descriptions of the spectral analysis and the dispersion analysis that follow assume the data sets are already in this format.

Spectral Analysis. Spectral analysis techniques provide a general way of characterizing the correlational structure of fluctuations in a series of successive response time measurements (Gilden et al., 1995; Gilden, 1997). There are several kinds of spectral techniques; what is referred to here as a spectral analysis is a particular method called power spectral density estimation, which yields a power spectrum of a trial series.

Successful applications of spectral methods require a certain amount of care, sophistication, and background knowledge. Accessible introductions to spectral techniques are provided by

Gottman (1981) and Chattfield (1996). Press, Teukolsky, Vetterling, and Flannery (1992) describe "how to" information as well as provide source computer code for the analyses. The information presented in Press et al. is rather technical but nevertheless very helpful. In fact, much of the information presented below is adapted from more formal treatments of the same topics in Press et al.

With the exception of the censored observations, each participant's trial series of time estimates is now ordered according to the trial on which the observation was collected—the order of the successive trials in the experiment. Connecting the points that represent the successive time estimates forms a complex waveform, as in Figure 6.1A. In a sense, a power spectral density analysis decomposes a trial series much as a prism breaks white light into its basic wavelengths, or colors. Spectral analyses decompose a trial series into a set of regular oscillations, component waves with particular frequencies and amplitudes. Taken together, the component waves mimic the overall pattern of oscillation in the observed trial series. An intuitive grasp of spectral analysis may be gleaned by thinking of it as a multiple regression analysis that fits a large set of simple sine (and/or cosine) waves to the complex response time waveform. The period of oscillation (the inverse of frequency) and the amplitude (relative height) of each component wave can vary. Oscillations corresponding to quickly changing trial-to-trial "jitter" map to a high frequency wave. Persistent excursions in one direction or the other from the mean of the trial series, over the course of, say, tens to hundreds of trials, map to lower and lower frequency oscillations. The output of a spectral analysis is a set of coefficients that characterize the

relative amplitudes of all the wave forms, ordered from lowest to highest frequency. This output is called the *power spectrum* of the signal. Loosely speaking, the spectral density coefficients correspond to a relative sum of squares for each frequency of sine wave that is used to fit the trial series. This is a bit like an r^2 for each sine wave that was passed through, or fit to, the series. Frequencies with larger amplitude coefficients imply more of the total variability was attributed to that particular frequency of oscillation.

As described, this basic recipe for a spectral decomposition procedure yields estimates of the amplitude (relative strength or energy) of many sinusoidal frequencies, but there is little statistical certainty in the magnitudes of any one of them. Each amplitude estimate is derived from just a single pass or "fit" of an individual ideal sine wave. Like any other statistical sample, the coefficient resulting from that fit can be unduly influenced by idiosyncratic properties of the data set in hand. Put differently, the standard deviation of each amplitude estimate for each particular frequency is huge—100% of its value (Press et al., 1992). Increasing the number of data points by using progressively longer data series only allows more and more frequencies to be estimated. A straight spectral density analysis always yields about half as many frequencies as there are data points (the highest resolvable frequency oscillates back and forth on every other data point). Thus, analyzing the entire data sequence at once yields maximum frequency resolution (many different sized sine waves are approximated), but does not lower the variability in the estimate of the amplitude for any particular frequency (see Press et al.). In terms of statistical certainty, the output is about as trustworthy as a factorial

ANOVA that has many, many experimental cells, but each cell mean is based on just one data point.

For data sets as variable as response time trial series, the procedures for computing a power spectrum must be adapted to balance the need to identify fluctuations over a suitably wide range of frequencies while simultaneously minimizing error variance in the estimation of the magnitude of any particular frequency. This is accomplished by breaking a single long 1024-trial series into several shorter, overlapping series of response times. This procedure is called data segmenting or data blocking. The individual power spectra derived from each short sub-series are then averaged. The averaging reduces the variability in the power spectral density estimate at each frequency; the cost is a reduction in the maximum number of frequencies that may be estimated.

Finally, the mathematics that govern the translation from the *time domain* (the trial-by-trial representation of the data) into the *frequency domain* (the frequency-by-frequency representation of the data) require the use of a procedure called *data windowing*. Essentially, a difficulty emerges from the fact that real data sets have distinct beginnings and ends, but the mathematics of spectral analysis assumes a data set of infinite length that is, more or less, strictly periodic. The consequence of this mismatch is a tendency for the power or energy associated with any particular frequency to "reverberate" or be blurred into the amplitude estimates of nearby frequencies. A typical data window applies a weighting function to the segment of the trial series that is undergoing spectral analysis. Data windows are designed to smooth the transition into and out of the data, and work something like

slowly turning up "volume" of the data starting at the beginning of the series, until it is at a maximum at the center of the series, and then slowly turning it down to zero again by the end of the sample (see Press et al., 1992 for details). The data blocking, spectral density averaging, and data windowing procedures, or similar statistical fixes that address the same issues, are often available as standard options in many spectral analysis computer routines.

If you are using a spectral density routine that returns the frequencies and power (the square of the absolute value of each amplitude) in linear units, the first step is to delete the highest (Nyquist) and lowest frequency (DC) coefficients, and then to transform the remaining coefficients by taking the log, base 10, of each frequency and its corresponding power estimate. Plot log-frequency against log-amplitude in a scatter plot. Reasonable evidence for inverse power-law scaling in the form of pink noise appears as a negatively sloped linear relation between the two variables on the log-log scatterplot. The linear relation must span a range of *at least* 2 decades of frequency (i.e., 2 log units, or 100 frequencies; Eke et al., 2000, 2002). Since natural fractals exhibit self-affinity across only a finite range of scales, the inverse power-law scaling relation may break down at either the highest or lowest frequencies, or both. It is also notable that the strength of the linear relation trades off with the number of frequencies that are used in the analysis. For a given series length, estimating the amplitudes of more frequencies typically yields a wider scatter of points in the log-log scatter plot. This is a consequence of the issues relating to the statistical certainty in the spectral coefficients, as was discussed previously.

The next step is to determine the value of the scaling exponent, the α in the scaling relation $1/f^{\alpha}$ (where f denotes frequency). The scaling exponent describes how the amplitude of the fluctuations change or scales as a function of their frequency. The easiest way to estimate α is to determine the slope of a least-squares regression of power as a function of frequency, using the logarithmically transformed values. (Technically, it is more correct to fit a least-squares power law in the linear domain but few researchers feel this extra step is critical). The slope of the regression line is the scaling exponent α. Response time series usually yield negatively accelerated slopes (recall that $1/f^{\alpha} = f^{-\alpha}$) that, within certain boundary conditions, discussed later, indicate pink noise, or slopes that are statistically equivalent to zero, which suggests white noise.

One additional difficulty with the spectral method is that the high frequency portion of the power spectrum can sometimes be "whitened," which appears either as flattening to zero slope at the highest frequencies or as combination of linear and U-shaped quadratic trends in the log-log regression that looks a bit like a tilted and mirror-reversed J. Figure 6.1C and Gilden et al.'s (1995) plots for the shorter time estimates display evidence of this pattern. The flattening at high frequencies may simply indicate a breakdown in the scaling relation at the highest frequencies, but in response time trial series it was linked to issues related to experimental design and measurement procedures that add sources of white noise to the signal (Gilden et al., 1995; Gilden, 1997; Gilden, 2001; Van Orden et al., 2003). For instance, very high frequency oscillations that unfold on a pace *faster* than the trial-by-trial pace of measurement may be "aliased" into the power of measured

frequencies. That tends to whiten the high-frequency end of a $1/f$ spectrum. *Aliased* frequencies refer to oscillations that live outside of the measured frequency range, and that are misinterpreted by the spectral analysis as different frequencies that fall within the measured range of frequencies (Press et al., 1992). It is similar to the way a person dancing in relative dark under a regularly flashing strobe light can be perceived as not moving, or to the apparent, but false, appearance of backward rotation of spoked wheels that one sometimes notices in old films. Excluding the highest frequencies in the log-log regression is often recommended to avoid the whitening of the high-frequency end (Eke et al., 2000, 2002).

The details of how the spectral analysis on the temporal estimates were conducted on the trial series appearing in Figure 6.1A can be summarized as follows. The spectral analyses resulted from averages of seven successive power spectra computations taken across successive sub-blocks of 256 trials. Each sub-block was multiplied by a triangular (Bartlett) window and the power spectrum was computed. The trial series was then shifted by 128 trials (1/2 the sub-block length) and a new power spectrum was computed. This process was repeated until the end of the series was reached. Thus, the power spectrum resulting from each participant's trial series was based on an average of seven (semi-independent) samples of the data set. This process yielded estimates of 129 (n/2 + 1) frequencies, but the highest and the lowest frequency were dropped, resulting in a total of 127 frequencies.

Figure 6.1B displays the results of the spectral analysis on linear scales. The x-axis depicts frequency, ranging from low to high. The y-axis depicts power, the square of the absolute value of each amplitude.

Figure 6.1C displays the results of same spectral analysis, now depicted on log-log scales. The approximately linear relation between the two variables in the log-log domain implies an inverse power-law scaling relation, consistent with pink noise. The slope of the regression line is –0.59, which corresponds to a $1/f^{0.59}$ scaling relation. Note, however, that the highest frequencies in Figure 6.1C seem to be slightly "whitened," which introduces a slight quadratic trend to the power spectrum and a slight bias in the slope of the regression line toward a shallower value. This pattern could justify eliminating the highest frequencies from the regression line by fitting only the lowest 25% of the frequencies, for instance. As expected, excluding the whitened higher frequencies yields a steeper spectral slope of –0.86.

As an additional check, it is important to recompute the spectral analysis using both more and fewer frequencies, and see comparable results. Analyses that use fewer frequencies should better resolve the linear nature of the scaling relation, and analyses using more frequencies should suggest that the scaling relation reaches into the lower frequencies, but it is critical to eliminate the detrending steps when examining the coefficients for lowest frequencies. The presented analysis, using 127 frequencies, reflects a compromise between the need for a satisfactory level of statistical certainty in the spectral coefficients (by examining frequencies that correspond to scales no larger than 1/4 the length of the series) and the need to establish the scaling range across at least 2 decades of frequencies.

Assuming a lack of evidence for white noise, the main reason for conducting a spectral analysis is to determine whether the value of the slope of the log-log regression line lies very near or less than –1, which

marks the boundary between a stationary fractional Gaussian noise and a nonstationary fractional Brownian motion. Dispersion analysis more accurately characterizes the fractal structure of a trial series than spectral analysis, but it can only be used on approximately or weakly stationary data sets. In this regard, idealized pink noise, with a spectral slope of –1, marks the boundary between mathematically stationary and nonstationary trial series. Fractional Brownian motions are nonstationary, and require other techniques, such as detrended fluctuation analysis (Peng, Havlin, Stanley, & Goldberger, 1995), which can be used on fractional Gaussian noise as well, or rescaled range analysis (Cannon et al., 1997). Those techniques are closely related to dispersion methods, but they will not be covered in this chapter. Response times often yield scaling exponents suggesting pink noise, and indicate that dispersion analysis is appropriate to determine the fractal dimension of the series. Next, the spectral slopes that define fractional Gaussian noise and fractional Brownian motions are described in detail.

Variability Categories. The magnitude of the power-law scaling exponent circumscribes at least two general classes of temporal variability that are called *fractional Gaussian noise* (fGn) and *fractional Brownian motion* (fBm) (see Mandelbrot & Wallis, 1969a/2000; Eke et al., 2000, 2002). Collectively, they are often referred to as $1/f^{\alpha}$ noise. The original theoretical development of these ideas used Gaussian-shaped probability density functions, but the classifications generalize to empirical data that have non-Gaussian density functions (Mandelbrot & Wallis, 1969c/2000).

Figure 6.3 illustrates how fGn and fBm can be described using the spectral slope of the power-law scaling relation. fGn exhibits log-log spectral slopes that range between 1 and –1. A slope of 0 indicates no historical dependence—independent sources of random variation, or *white noise*. Slopes reliably greater than 0 and less than 1 indicate a tendency for positive data values to be followed by negative values, which is termed *anti-persistence*; this is sometimes called *blue noise*. Slopes less than 0 and greater than –1 indicate *persistence*—positive data values tend to be followed by positive data values. This is the domain of *pink noise*, the main topic of this chapter. Assuming the log-log regression is linear across 2 or more decades (log units) of frequency, pink noise is simply a statistically reliable departure from white noise in the direction of persistence, evaluated using a combination of spectral and fractal analyses.

Spectral slopes less than –1 and greater than –3 describe a related but fundamentally different kind of variability called fractional Brownian motion. This is the domain of *random walks*. A spectral slope of –2 indicates idealized Brownian motion. Slopes less than –1 but greater than –2 indicate anti-persistent fractional Brownian motion, in which successive increments tend to have opposite signs. Slopes between –2 and –3 indicate persistent fractional Brownian motion, in which successive increments tend to have the same sign (this is sometimes called *black noise*). Notably, a single parameter characterizes this entire family of noises.

Idealized pink noise, or $1/f^1$ noise, is special mainly because when the x- and y-axes of a $1/f^1$ noise are enlarged in like proportions; the enlarged portion of the series is statistically indistinguishable from

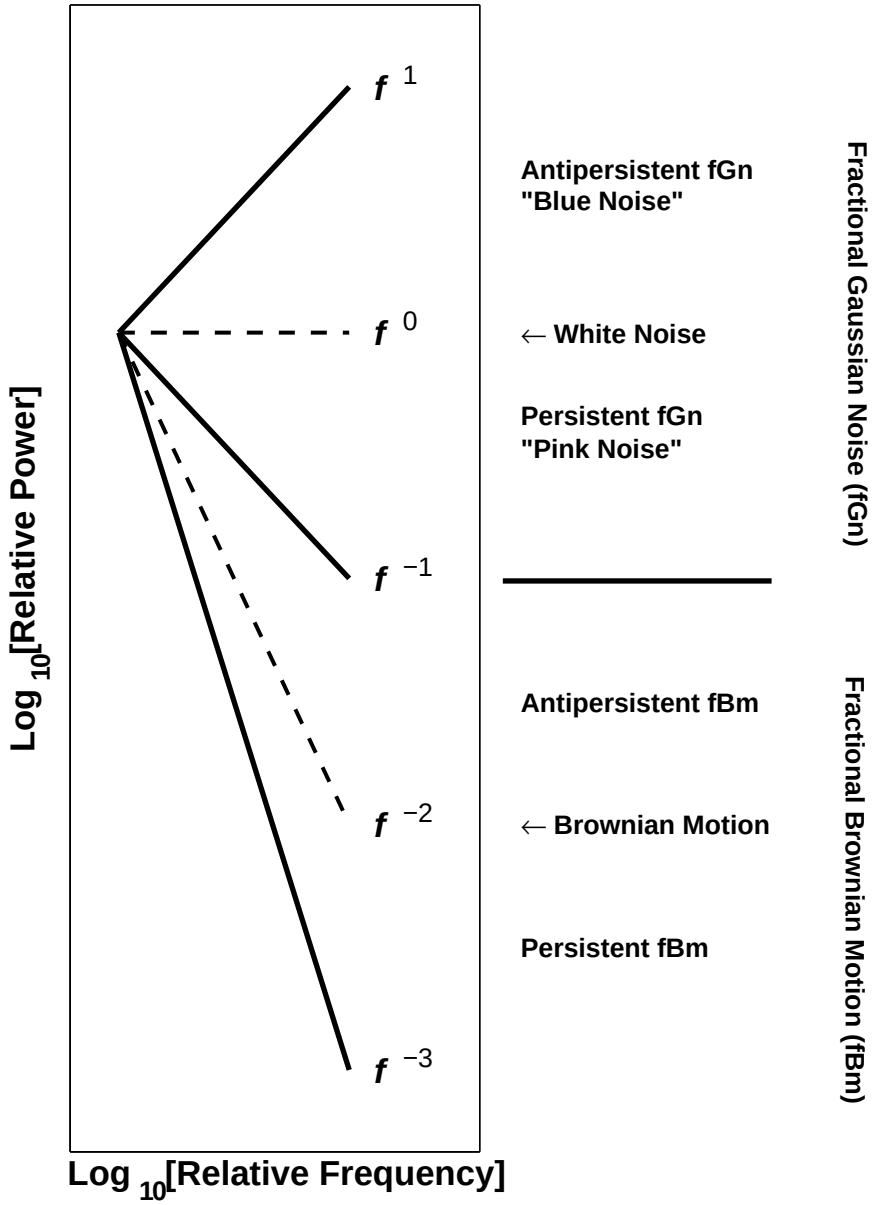

Figure 6.3. The figure adapted from Eke et al. (2000) portrays an idealization of spectral slopes that distinguish fractional Gaussian noises (fGn) and fractional Brownian motions (fBm) (see also Cannon et al., 1997). Idealized pink noise, or $1/f^1$ noise, is special as a mathematical way-point; it marks the boundary between stationary (fGn) and nonstationary (fBm) data series, two categorically distinct kinds of variability (Eke et al., 2000, 2002).

the original series. Pink noise that has a scaling exponent that falls between 0 and 1 requires the x- and y-axes to be enlarged in *different* proportions to yield the same effect (Eke et al., 2000, 2002). Idealized pink noise is also important as a mathematical way-point, as it marks the analytic boundary between stationary and nonstationary data series. Idealized pink noise, and data series with scaling spectral slopes less than –1, have in theory, and in practice, *infinite variance*. Pink noise with nontrivial spectral slopes less than –1 has infinite variance *in practice*, in the sense that the interdependence of finite statistical samples yields unreliable population parameters (Bassingthwaighte et al., 1994; Mandelbrot & Wallis 1969b/2000). Thus, a spectral slope of –1 marks an important boundary between two categorically distinct kinds of variability, fGn and fBm (Eke et al., 2000, 2002).

Many natural systems emit pink noise, but spectral slopes of exactly –1 are not usually observed. Heart rate variability can exhibit a spectral slope very near –1 (e.g., Eke et. al., 2002), but many established examples display slopes between 0 and –1. A large sample of yearly tree ring indices has average spectral slopes of –0.43. Annual precipitation statistics have average slopes of –0.48. The classic Nile River yearly minimum series yields a spectral slope of –0.82 while measurements of the Nile's yearly maximum levels display a slope of about –0.68 (Mandelbrot & Wallis, 1969b/2002). Natural phenomena entail sources of unsystematic external variability, in addition to sources of intrinsic $1/f$ scaling, which results naturally in pink noise with spectral slopes greater than –1.

Typically, response times from elementary cognitive tasks, such as temporal estimation, yield scaling exponents that fall within the range of 0-1, and thus contain a nested, statistically self-similar pattern of positive correlation across successive observations. Spectral slopes that lie near –1 suggest that the nested structure of positive correlation dominates the series. Spectral slopes that lie closer to zero indicate a less prominent, "whitened" structure of nested positive correlation across the series.

Next, a statistical technique called dispersion analysis is introduced. It more accurately characterizes the pattern of statistical self-similarity than spectral analysis. It yields a fractal dimension statistic (FD), which is closely related to the slope of the power spectrum. In fact, FD = 1 + (S + 1)/2, where S is the spectral slope, of the log-log regression line. This formula assumes the spectral slope of the series falls between –1 and 0. Note, however, that since the two analyses "break up" the trial series in mathematically different ways (e.g., Fourier analysis versus means and standard deviations) they will not typically output exactly the same fractal dimension for the exact same signal, although they should yield reasonably similar outcomes. If there is a strong disagreement between the two methods, examine the signal, and the steps in the analysis, carefully for potential artifacts.

Dispersion Analysis. Dispersion analysis yields the fractal dimension of a trial series and gauges the change in variability due to changing sample sizes. Dispersion analysis determines whether the trial series variability converges fast enough, as sample size increases, to yield stable statistical estimates of population parameters. If not, then the process that produced the variability is, in practical terms,

scale free in the sense that it has no characteristic "quantity" or scale of variability.

There are several ways to compute the fractal dimension, and dispersion techniques are among the most accurate (Bassingthwaighte et al., 1994; Caccia et al., 1997; Eke et al., 2000, 2002). Spectral analyses yield less reliable fractal dimension estimates than dispersion methods. A practical advantage of dispersion analysis is that familiar statistical constructs, means and standard deviations, are used for the analysis. A version of the standard technique of *relative dispersion analysis* is presented here. It allows for the use of normalized data instead of raw data—call this *standardized dispersion analysis* to avoid confusion with other methods. Note that the standardized dispersion analysis yields dispersion measurements that are in units of the standard error of the mean; the standard deviation of a sampling distribution of means, comprised of means of (adjacent) samples of specified sizes. By contrast, the original method described by Bassingthwaighte et al. is based on the relative dispersion statistic that is comprised of a ratio of the standard deviation and the mean (i.e. RD = SD/M). The outcomes of the two techniques are identical. However, the spectral technique presented earlier assumes a normalized trial-series and standardized dispersion analysis allows the same detrended and normalized data set to be submitted to both analyses.

When computing the dispersion statistics in the subsequent steps, compute the standard deviation using the population formula (i.e. use N, the number of data points, in the calculation, rather than the usual bias corrected $N - 1$).

A dispersion analysis repeatedly resamples the trial series using sampling units of different sizes to estimate the fractal dimension of a trial-series. In the steps that follow, variability is gauged using the standard deviation of means of progressively larger adjacent samples. That is, the analysis tracks how variability in sample means decrease as progressively larger samples of adjacent data points are aggregated together in a sample mean. If the samples are statistically independent, then it should not matter that adjacent samples are being grouped and regrouped to form samples of different sizes.

To perform the analysis first construct a table, like Table 6.1. Begin the table by recording a 1 in the points-per-bin column and another 1 in the dispersion column. The overall standard deviation (SD = 1) of the normalized series represents the overall dispersion of the series, given that the data points are treated individually. Essentially, the overall trial series standard deviation is treated as a population parameter, and for this initial step, N is also 1. The overall standard deviation is identical to the variability of a sampling distribution of 1024 "means," computed across single, individual observations, which is just the raw score standard deviation.

The next step involves grouping the data points into adjacent pairs, which makes it more obvious that variability in the means of sample bins is being tracked. Compute the mean for each successive pair of points—*each bin*. This yields the new set of data points; a sampling distribution of means that contains the 512 values of each 2-point mean. Compute the standard deviation of this new distribution. Enter a 2 in the points-per-bin column of the table, because 2 point sample means were used. Then enter the standard deviation of the

sample means in the dispersion column. If the trial series were composed of statistically independent observations, then the expected standard deviation of the sampling distribution of the 512 two-point means would be $1/\sqrt{2}$ or about 0.71. Interested readers may consult Van Orden et al. (2003) for an explanation of the relation between the standardized dispersion statistic and the equation for the standard error of the mean.

Table 6.1: Standardized dispersion as a fucntion of sample bin size in linear and logrithmic units.

Bin Size	**Standardized Dispersion**	**Log_{10} (Bin Size)**	**Log_{10} (Standardized Dispersion)**
1	1	0	0
2	0.88	0.30	–0.06
4	0.79	0.60	–0.10
8	0.73	0.90	–0.14
16	0.68	1.20	–0.17
32	0.60	1.51	–0.22
64	0.54	1.81	–0.27
128	0.47	2.11	–0.32
256	0.15	2.41	–0.82
512	0.02	2.71	-1.68

Repeat the previous step until only two data points are left (i.e., the third iteration will use 256 bins of every four successive data points, the fourth iteration uses 128 bins of size eight, and so on, until there are two bins of size 512). At the culmination of each step, enter the number

of points that comprises each bin, and the standard deviation of the distribution of the sample means into the table. In the final repetition, the final two data points come from a bin containing the first half of the original trial series and a bin containing the last half. To summarize: Each step in constructing the table generates an *N* that is equal to the bin size and a standard deviation that estimates dispersion at that bin size. If the trial series was a series of statistically independent data points, then the standard deviation should diminish very nearly as a function of $1/\sqrt{N}$, as the size of the sample bin sizes are progressively increased.

Finally, plot the logarithm of the numbers in the points-per-bin column against the logarithm of the numbers in the standard deviation column as in Figure 6.2. Base-10 logarithms were used here, but other bases also work. For instance, using base 2 represents the number of samples in the bins as integer powers of 2—just be sure to use the same base for taking the log of the dispersion values as well. The relation between the two variables should be linear on double-log scales, except perhaps for the three or four points that correspond to the largest bins. Typically, the last few relative dispersion measurements that correspond to the very largest bin sizes are excluded at this point (Cannon et al., 1997); this is a critical adjustment when detrending is used. Here, the three largest bins were excluded because the detrending procedures removed the variability at these scales. Natural fractals exhibit scaling relations across a finite range of scales, so the linear relation is expected to break down at some point, for either (or both) the smallest or largest bin sizes. (Points excluded in the log-log regression in Figure 6.2 appear as open circles.) If using a

standardized series the dispersion values at the largest bin size approach zero (and negative infinity when the log transformation is performed). As such, they bias the slope of the regression line (see Caccia et al., 1997, for additional refinements of this technique, especially for shorter data sets).

A linear relation with a negative slope in log-log coordinates establishes an inverse power-law scaling relation and indicates that the trial series is a simple fractal (Bassingwaighte et al., 1994). The fractal dimension of the series is given by subtracting the slope of the least-squares regression line from one, the Euclidean dimension of the series. The relation illustrated in Figure 6.4 is an inverse power-law scaling relation. The slope of the log-log regression line is –0.14, and the fractal dimension of this trial series is therefore 1.14. Transforming the spectral slope of –0.59 into a fractal dimension yields 1.21; although those values are not identical, they are in the same neighborhood. Also recall that the spectral plot revealed evidence that the high-frequencies were "whitened," which tends to bias the spectral slope towards shallower slopes, and thus, larger fractal dimensions. For instance, returning to the spectral coefficients and fitting only the lowest 25% of the power spectrum coefficients yields a spectral slope of –0.86, which translates to a FD of 1.07.

At this point, a reader might ask him or herself, which characterization is the correct one? It is important to be mindful of the fact that each method of analysis has strengths and susceptibilities. For example, the spectral methods are sensitive to a host of artifacts that affect the high-end of the frequency range, but dispersion analysis is not as susceptible to these influences since it is not based on a Fourier

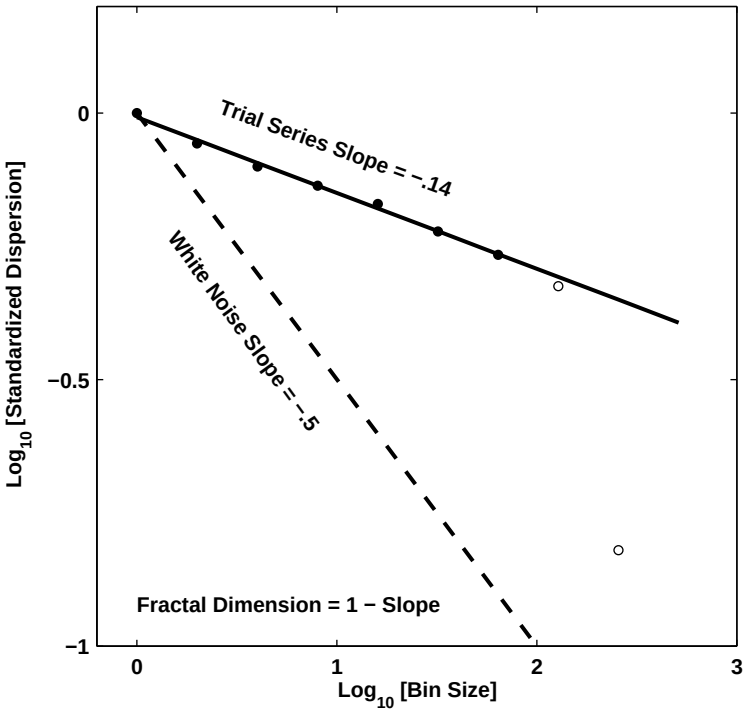

Figure 6.4. Standardized dispersion is depicted as a function of sample-bin size, on double logarithmic scales. The x-axis indexes the base-10 logarithm of the number of adjacent data points in each adjacent sample bin. The y-axis indicates the base-10 logarithm of the standard deviation of the mean standardized dispersion measures, across all the sample bins. The solid line is a least-squares regression line for the first 7 data points, represented by the solid points. The three points depicted as open circles correspond to the three largest bins and were not included in the regression analysis because the detrending procedures tends to eliminate variability at those larger scales. The open circle corresponding to the 2 largest 512 point bins is not shown because it fell below the limit of the y-axis.

transform. Moreover, the earlier discussion of spectral analysis illustrates that the raw data is subjected to a series of transformations to yield a spectral slope. Dispersion analysis, on the other hand, can be unduly influenced by simple linear trends that span the full length of the series. Dispersion analysis has been shown to lose some of its accuracy in characterizing time series that display very strong

interdependence, such as pink noise with a spectral slope near –1, (Eke et al., 2000, 2002), which are essentially nonstationary signals in the case of limited sample sizes. Dispersion analysis should not be used on fBm series, which are truly nonstationary signals.

Most importantly, both methods should yield solid, converging evidence of the presence of a power-law scaling relation. All other things being equal, the fractal dimension based on the dispersion method is more accurate, but it would not be unreasonable to report an average across these two, or perhaps additional methods in research reports. Notice the complementary nature of the two analyses; spectral methods are reliable at the intermediate and larger scales (the intermediate and lower frequencies), while dispersion methods are reliable at the intermediate and smaller scales (medium and higher frequencies).

The take-home point is that both spectral and dispersion methods decompose the raw data in different ways, and thus interact with the idiosyncrasies of a given empirical signal in slightly different ways. In addition, both analyses require the researcher to make a number of choices about a range of parameters, such as the manner of detrending, the number of spectral coefficients or bin sizes to fit with a regression line, and so on. Each choice will impact the outcome of the analysis in some way, at least compared to other choices that could have been made. As such, both methods should agree qualitatively, and yield similar, but not necessarily identical, fractal dimension estimates. Redoing the analysis using other parameter choices may change the value of the fractal dimension statistics somewhat, but it will typically do so in a systematic manner (e.g., fitting the whole spectrum

often yields shallower spectral slopes). As long as the fractal description survives trivial changes in the parameters that are used to govern the analysis (e.g., dropping the 3 or 4 largest bins in the dispersion analysis or fitting the entire spectrum versus the lowest 25%-30% of the spectrum), it is likely accurate. It is then up to the researcher to present a succinct, conservative analyses that, in his or her best judgment, accurately portrays the data at hand.

Significance Testing. There are two general kinds of statistical tests that one may want to conduct on the fractal dimension or spectral slopes resulting from the analyses discussed earlier. The first situation involves experimental research designs that entail contrasts between two or more conditions, and standard linear statistical methods such as a t-test or ANOVA (or nonparametric equivalents) can be used to establish differences in fractal dimension across groups. However, such comparisons must be carefully considered. Simple issues of measurement affect the fractal dimension and spectral slope of trial series arising from different tasks (e.g., Van Orden et al., 2003), so the fact of a statistically significant difference across tasks may not point to a theoretically interesting difference.

The second, and perhaps more novel, approach is to determine whether white noise or pink noise better describes an observed trial series. There are a number of ways that one or several trial series can be statistically distinguished from white noise. If a reasonably sized distribution of slope or fractal dimension estimates is available from several subjects, a single sample t-test could be used to determine if the sample mean of the spectral slopes is significantly different from 0, the expected slope of white noise (or if the fractal dimension differs

from 1.5). If only a single trial series is available, then one can easily test for a significant departure from a pattern of white noise by contrasting the observed data with surrogate data sets (e.g., the reshuffled data in Figure 6.1D.). Simply randomly reshuffle the order of the original trial-series, 10 or 20 times, each time computing the fractal dimension and/or spectral slope of each reshuffled version of the trial-series. Then calculate the fractal dimension mean and SD of the surrogate data sets and compare the mean to the observed values of the original series. If the fractal dimension and spectral slope of original series is in the pink noise range, and more than 3 SD away from the mean of the surrogates, then a null hypothesis of white noise can be rejected (Hausdorff, 1996; see Theiler et al., 1992, and Efron & Tibshirani, 1993, for more details and options).

DISCUSSION

At first glance, Gilden et al.'s (1995) original temporal estimation task appears to be directed at understanding a person's ability to replicate a range of fixed time intervals. Thus, one might expect it to include an extensive discussion of how well or poorly people performed, but accuracy was barely mentioned in their report. The focus was instead on the structure of intrinsic patterns of trial-by-trial variability.

Conventional research methods in psychology emphasize a "cause-effect" metaphor as the only sensible route to scientific understanding. Experimental manipulations are introduced as a way of brining empirical phenomena under experimental control. Thus,

factorial manipulations are designed to control empirical "effects." Perhaps one experiment is designed to amplify an experimental effect, and another design may diminish the same effect. All hope of a plausible scientific explanation hinges on reducing experimental manipulations to causal factors inside the mind or brain (Van Orden, Holden, Podgornik, & Aitchison, 1999). As such, one impulse that cognitive psychologists sometimes have, especially those with solid training in conventional statistics, is to think of pink noise in terms of a statistical nuisance. If pink noise causes problems for inferential statistics (and, therefore, for the aforementioned causal logic), then perhaps pink noise can be removed, modeled, or otherwise controlled. This is a natural move for one taking a purely statistical perspective on data. However, finding pink noise can lead to fundamentally different scientific perspectives on the data. For example, one may wonder how the human mind and body might be organized to give rise to statistical interdependence, and to interact across different time scales. From a scientific perspective it seems to make sense to ask what the empirical fact of pink noise in cognitive and other human performance implies about the fundamental coupling of bodily processes.

For cognitive science, Gilden's approach was revolutionary because it supplied the seed of an alternative format for scientific explanations of cognitive performance, and for psychology in general. That alternative began with attempts to establish the presence of pink noise *despite* a range experimental manipulations (e.g., Gilden et al., 1995), and it was subsequently expanded to include experiments in which pink noise survived changes in cognitive tasks (e.g., Cayton & Frey, 1997; Gilden, 1997, 2001; Van Orden & Holden, 2002; Van Orden

et al., 2003). Thus, the alternative approach was directed at the identification of ubiquitous performance phenomena, in the hope that those phenomena may illustrate something very general about human performance. Gilden was perhaps one of the first to suggest, on the basis of an empirical finding, that pink noise implied that complex systems theory may be relevant to cognitive psychology (Gilden et al., 1995). The link is implied, in part, because pink noise is an important (but not sufficient) footprint of self-organization and self-organizing systems in nature (see also Aks, Chapter 7, and Van Orden et al., 2003). More generally, the widespread finding of pink noise in trial series from exhaustively studied, standard cognitive tasks underscores the potential for analyses of patterns of *variability* in cognitive performance to be informative.

Author Note

I gratefully acknowledge Vahe Amirian for his help in collecting the data for the timing experiment, and the helpful comments and feedback I received from three reviewers while preparing this chapter.

References

Aks, D. J., Zelinsky, G. J., & Sprott, J. C. (2002). Memory across eye movements: 1/f dynamic in visual search. *Nonlinear Dynamics, Psychology, and Life Sciences, 6*, 1-25.

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

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.

Cannon, M. J., Percival, D. B., Caccia, D. C., Raymond, G. M., & Bassingthwaighte, J. B. (1997). Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A, 241*, 606-626.

Chatfield, C. (1996). *The analysis of time series: An introduction* (5th ed.). New York: Chapman & Hall.

Clayton, K., & Frey, B. B. (1997). Studies of mental "noise". *Nonlinear Dynamics, Psychology, and Life Sciences, 1*, 173-180.

Delingières, D. Fortes, M., & Ninot, G. (in press). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology, and Life Sciences*.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Eke, A., Hermán, P., Bassingthwaighte, J. B., Raymond, G. M., Percival, D. B., Cannon, M., Balla, I., & 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.

Feder, J. (1988). *Fractals*. New York: Plenum Press.

Forster, K. I., & Forster, J. C. (1996). *DMASTR Display System for Mental Chronometry (Version 5.18)* [Computer Software]. Tucson, AZ: Authors.

Gilden, D. L. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science, 8,* 296-301.

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

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science, 267*, 1837-1839.

Gottman, J. M. (1981). *Time-series analysis : A comprehensive introduction for social scientists*. Cambridge, UK: Cambridge University Press.

Hausdorff, J. M., Purdon, P. L., Peng, C.-K., Ladin, Z., Wei, J. Y., & Goldberger, A. L. (1996). Fractal dynamics of human gait: Stability of long-range correlations in stride interval fluctuations. *Journal of Applied Physiology, 80*, 1448-1457.

Jensen, H. J. (1998). *Self-organized criticality*. Cambridge, UK: Cambridge University Press.

Kelly, A., Heathcote, A., Heath, R., & Longstaff, M. (2001). Response time dynamics: Evidence for linear and low-dimensional structure in human choice sequences. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 54A*, 805-840.

Keppel, G., & Zedeck, S. (1986). *Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches*. New York: W. H. Freeman & Co.

Mandelbrot, B. B. (1982). The fractal geometry of nature. San Francisco: W. H. Freeman & Co.

Mandelbrot, B. B., & Wallis, J. R. (1969a). Computer experiments with fractional Gaussian noises, Part 1, 2 & 3. Water resources research, 5, 228-267. Reprinted in Mandelbrot B. B. (2000), *Gaussian Self-Affinity and Fractals: Globality, the Earth, 1/f Noise, and R/S*. New York: Springer-Verlag.

Mandelbrot, B. B., & Wallis, J. R. (1969b). Some long run properties of geophysical records. Water resources research, 5, 321-340. Reprinted in Mandelbrot B. B. (2000). In *Gaussian Self-Affinity and Fractals: Globality, the Earth, 1/f Noise, and R/S*. New York: Springer-Verlag.

Mandelbrot, B. B., & Wallis, J. R. (1969c). Robustness of R/S in measuring noncyclic global statistical dependence. Water resources research, 5, 967-988. Reprinted in Mandelbrot B. B. (2000). In *Gaussian Self-Affinity and Fractals: Globality, the Earth, 1/f Noise, and R/S*. New York: Springer-Verlag.

Peitgen, H.-O., Jürgens, H., & Saupe, D. (1992). *Chaos and fractals: New frontiers of science*. New York: Springer-Verlag.

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipies in C* (2nd ed.). Cambridge, UK: Cambridge University Press.

Riley, M. A., Wong, S., Mitra, S. & Turvey, M. T. (1997). Common effects of touch and vision on postural parameters. *Experimental Brain Research, 117*, 165-170.

Schroeder, M. R. (1991). *Fractals, chaos, power laws: Minutes from an infinite universe*. New York : W.H. Freeman.

Theiler, J., Eubank, S., Longtin, A., Gladrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: The method of surrogate data. *Physica D, 58*, 77-94.

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., Podgornik, M. N., & Aitchison, C. S. (1999). What swimming says about reading: Coordination, context, and homophone errors. *Ecological Psychology, 11*, 45-79.

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