

THEORETICAL AND REVIEW ARTICLES

Estimation and interpretation of $1/f^\alpha$ noise in human cognition

ERIC-JAN WAGENMAKERS

University of Amsterdam, Amsterdam, The Netherlands

SIMON FARRELL

University of Bristol, Bristol, England

and

ROGER RATCLIFF

Ohio State University, Columbus, Ohio

Recent analyses of serial correlations in cognitive tasks have provided preliminary evidence of the presence of a particular form of long-range serial dependence known as $1/f$ noise. It has been argued that long-range dependence has been largely ignored in mainstream cognitive psychology even though it accounts for a substantial proportion of variability in behavior (see, e.g., Gilden, 1997, 2001). In this article, we discuss the defining characteristics of long-range dependence and argue that claims about its presence need to be evaluated by testing against the alternative hypothesis of short-range dependence. For the data from three experiments, we accomplish such tests with autoregressive fractionally integrated moving-average time series modeling. We find that long-range serial dependence in these experiments can be explained by any of several mechanisms, including mixtures of a small number of short-range processes.

In the vast majority of experiments in cognitive psychology, any information about the order in which empirical observations are obtained is discarded. The implicit assumption of this approach is that performance on successive trials is not correlated and, hence, any variance unaccounted for by the experimental manipulations can be considered random noise. In this article, we study human performance in cognitive tasks from a quite different perspective, focusing on how performance fluctuates over time (cf. Gilden, 2001). In this time-series approach to human cognition, the order in which observations are obtained is not neglected but, in fact, provides the data of interest. Thus, we seek to study the extent to which current

behavior can be explained on the basis of past behavior (or, alternatively, the extent to which we can predict future behavior). In addition, the specific manner in which serial correlations decrease as the number of intervening trials increases can provide information about underlying cognitive mechanisms.

Several researchers in cognitive psychology have studied serial correlations in human performance (e.g., Merrill & Bennett, 1956; Verplanck, Collier, & Cotton, 1952; Weiss, Coleman, & Green, 1955). In particular, Laming (1968) reported and modeled the serial dependency structure of response times (RTs) in two-choice experiments. Laming (1968) found that RTs on contiguous trials were positively correlated, so that a relatively short RT tended to be followed by another relatively short RT. These serial correlations decayed with increasing separation between trials (i.e., *lag*), so that there was evidence of significant serial correlations only up to a lag of six or seven. The fast decay of these serial correlations is consistent with other work (e.g., Laming, 1979a) that has found serial correlations to be small and transient (or even nonexistent—e.g., Busey & Townsend, 2001; Townsend, Hu, & Kadlec, 1988).

However, the popular belief that serial correlations in human performance are small and transient has recently been challenged. Specifically, it has been reported that the pattern of serial correlations in human performance

Preparation of this article was supported by NIMH Grant R37-MH44640, NIA Grant R01-AG17083, and NIMH Grant K05-MH01891. We thank David Gilden for helpful discussions and for sending us the data from his spatial estimation experiment (Gilden, 2001, p. 48). We thank Andrew Heathcote, David Gilden, Han van der Maas, Jerry Busemeyer, Manolo Perea, Marius Usher, and Robert Nosofsky for helpful comments on an earlier draft of this paper. Our data can be obtained from <http://www.psych.nwu.edu/~ej/fnoise/noisedat.html>. Correspondence concerning this article should be addressed to E.-J. Wagenmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, Amsterdam 1018 WB, The Netherlands, or to S. Farrell, Department of Experimental Psychology, University of Bristol, 8 Woodland Rd., Bristol BS8 1TN, England (e-mail: ewagenmakers@fmg.uva.nl or simon.farrell@bristol.ac.uk).

is characterized by a particular type of serial dependence known as $1/f$ noise (see, e.g., Chen, Ding, & Kelso, 1997; Gilden, 1997, 2001; Gilden, Thornton, & Mallon, 1995; Gilden & Wilson, 1995a; Gottschalk, Bauer, & Whybrow, 1995; Van Orden, Moreno, & Holden, 2003; Ward, 2002). Such $1/f$ noise has several defining properties that make it of considerable theoretical and practical interest. For example, $1/f$ noise is found to be self-similar in that it possesses the same temporal characteristics regardless of the time scale on which it is examined (this property is discussed in more detail later on). The property of $1/f$ noise that is most relevant for the present discussion is that the associated serial correlations decay very slowly as the number of intervening trials increases. The presence of $1/f$ noise therefore indicates that serial correlations are persistent, in contrast with the traditional view that they are transient. Transient serial correlations are said to be short-range dependent, whereas persistent serial correlations are said to be long-range dependent. The differences and similarities between short-range dependence (SRD) and long-range dependence (LRD, or $1/f$ noise) are the focus of this article.

Processes that show LRD have generated a tremendous amount of interdisciplinary scientific interest. One of the reasons for this is that LRD appears to be ubiquitous, yet it is not easily explained. Time series that show LRD have been observed in a wide variety of systems encompassing many different fields, such as physics, biology, technology, hydrology, geophysics, economics, and sociology. Specific examples that illustrate the ubiquity of LRD include the flow of the River Nile, stock market fluctuations, sunspot activity, neuronal spike trains, network traffic, pressure variations in music and speech, and magnetoencephalographic time series (see, e.g., Baillie, 1996; Beran, 1992, 1994; Davidsen & Schuster, 2000; Hausdorff & Peng, 1996; Kaulakys & Meskauskas, 1998; Novikov, Novikov, Shannahoff-Khalsa, Schwartz, & Wright, 1997; Pilgram & Kaplan, 1998; Schroeder, 1991; Voss, 1988; and references therein).¹

LRD has also recently been reported across a range of tasks in cognitive psychology. These tasks include lexical decision, serial and parallel visual search, mental rotation, simple RT, shape and color discrimination, temporal estimation, estimation of applied force, estimation of angular rotation, and estimation of spatial intervals (see, e.g., Chen et al., 1997; Gilden, 1997, 2001; Gilden et al., 1995; Gilden & Wilson, 1995a; Van Orden et al., 2003; Ward, 2002). It should be noted that the serial correlations observed in these studies are not only persistent (i.e., they decrease slowly) but also relatively large in absolute magnitude, exerting a powerful effect on task performance. For instance, Gilden (2001) demonstrated by analysis of his experiments that the presence of LRD accounted for a considerable proportion of variability in performance measures such as RT. The proportion of variability accounted for by LRD was sometimes greater than that accounted for by most standard manipulations in cognitive psychology (see, e.g., Gilden, 2001, p. 39).

On the basis of the ubiquity of LRD in cognitive psychology and a range of other fields, and on the basis of the considerable impact of LRD on task performance, it has been argued (e.g., Gilden, 2001; Van Orden et al., 2003) that the persistence of serial correlations sheds new light on the dynamics of human cognition, since it pertains to components of processing that are not addressed in current models of information processing (see, e.g., Gilden, 2001; Ward, 2002). Clearly, the fact that performance on trial n is dependent on performance on, say, trial $n - 30$ is puzzling and presents a novel challenge to current models of human information processing.

As we will show later, the distinction between SRD and LRD can be quite subtle. Therefore, it is especially important that systematic experimental investigations be carried out, accompanied by rigorous data analysis techniques. In addition, the presence of LRD in human performance needs to be explained in terms of underlying causal mechanisms. However, to date the research on $1/f$ noise and LRD has been rather exploratory in nature, and several areas for improvement can be identified.

One area for development is the method of identification of $1/f$ noise in psychological time series. In order to test for the presence of LRD in psychological studies such as choice RT experiments, the observed trial-by-trial sequence of RTs from a participant is kept intact² and the time series is usually subjected to analysis in the frequency domain. As will be explained in more detail later, such an analysis decomposes the time series into waves of different frequencies, each wave possessing a particular strength. The presence of LRD is often determined solely on the basis of such a frequency domain analysis, by judging the fit of a straight line through the average log-log power spectrum by eye. In psychological research, no principled statistical testing of LRD against the alternative explanation of short-term dependence is carried out.

A second area for theoretical development concerns what the presence or absence of LRD can tell us about human cognition. To date, no general explanation of the universality of LRD (e.g., Kaulakys & Meskauskas, 1998; Milotti, 1995) has been widely accepted. In the psychological literature, a thorough discussion of the various ways through which LRD can arise has not been attempted. One reason for this may be that it is as yet unclear under what specific conditions LRD occurs in psychological experiments, as the reported findings leave open some important questions (a review of the empirical findings will be presented later). Furthermore, few psychologically interpretable models that generate LRD have been proposed, underscoring the fact that psychological research for long-range processes is at this point phenomenon driven rather than theory driven (cf. Newell, 1973).

The organization of this article roughly follows the presentation of the issues above. First, we will discuss the defining characteristics of LRD. LRD is also known as *long memory*, *long-term correlations*, *strong dependence*, and *infinite memory*; we choose to use the present term to avoid confusion with psychological concepts

such as long-term memory. The discussion of defining characteristics will provide the necessary background knowledge for the later sections. We also point out that LRD and nonlinearity, which are sometimes associated in discussions of LRD, are independent concepts. In the subsequent sections, we discuss how rigorous identification of both SRD and LRD can be accomplished through autoregressive (AR) fractionally integrated moving-average (ARFIMA) time series modeling in combination with AIC model selection. With this approach, the fit of a model containing only short-term dependencies is compared with that of the same model with an additional LRD component. Monte Carlo simulations illustrate that, whereas the traditional spectral slope method is severely biased in the presence of SRD, the combination of ARFIMA modeling and AIC model selection is relatively robust to SRD.

After dealing with these methodological concerns, we report three experimental tasks designed to address several open questions that we encountered when surveying the psychological literature on $1/f$ noise. The results show that serial dependence is most pronounced in a temporal estimation task (cf. Gilden, 2001), and is at least as prominent in a simple RT task as it is in a choice RT task, in contradiction to previous claims (cf. Gilden et al., 1995). In all three tasks (i.e., simple RT, choice RT, and temporal estimation), ARFIMA analyses and accompanying model comparison techniques show support for the presence of LRD.

We then discuss various psychological models we claim can account for LRD, and we support these claims by ARFIMA analyses of the time series generated by the models. We point to general characteristics, such as aggregation of short-range processes, that may explain the phenomenon of LRD. For the temporal estimation task, we introduce a simple regime-switching model that is based on the assumption that subjects use different strategies throughout the experiment, implemented via discrete jumps of a temporal criterion. We also demonstrate how sequential sampling models (see, e.g., Laming, 1968, 1979a; Ratcliff & Rouder, 1998) could naturally incorporate similar mechanisms to account for long-range serial dependence in choice RT tasks.

WHAT IS $1/f^\alpha$ NOISE?

For a full understanding of the methods and arguments presented in this article, an elementary knowledge of time series analysis is required. We will first introduce some general properties of several time series models (Figures 1–5) and then present the defining characteristics of long-range processes in more detail.

Serial Dependence in Standard Time Series

Figure 1 shows prototypical time series generated by five different statistical models. The first model (Figure 1A) is a “purely random process” (Priestley, 1981), in

which successive observations are independently drawn from some distribution. Hence,

$$X_t = \varepsilon_t, \quad (1)$$

where X_t denotes the value of the observation on time t and ε denotes a randomly drawn value from a Gaussian distribution, for instance. (For the time series from Figure 1, ε is always drawn from a standard normal distribution.) The random variable ε is sometimes termed *innovation* in the time series literature. Thus, X_t forms a sequence of uncorrelated random variables—that is, for all $s \neq t$, $\text{cov}\{X_s, X_t\} = 0$; this model is said to generate *white noise*.

The second time series (Figure 1B) is generated by a *random walk process*. A random walk process is a running sum of independent observations—that is,

$$X_t = \sum_{s=1}^t \varepsilon_s \quad (2)$$

or, equivalently,

$$X_t = X_{t-1} + \varepsilon_t. \quad (3)$$

In contrast to the other time series discussed here, the random walk process is nonstationary because its variance increases over time. Equations 2 and 3 show that adding (integrating) successive observations generated by the purely random process yields a random walk process, and taking the difference between successive observations (i.e., differencing) in a random walk process yields a purely random process (see, e.g., Kasdin, 1995). Henceforth, the term *differencing* will refer to the procedure of constructing a new time series by the subtraction of the successive observations from the old time series—that is, by calculating the differences between X_t and X_{t-1} .

The third time series (Figure 1C) is generated by an autoregressive (AR) model. In an AR model, the value of the current observation depends partly on the value of the previous observation,

$$X_t = \phi_1 X_{t-1} + \varepsilon_t, \quad (4)$$

where the magnitude of the dependence is quantified by ϕ_1 , where $\phi_1 \in (-1, 1)$. For the present illustration, we let the value of an observation on time t depend only on the value of the observation on time $t-1$. This model is called a *first-order AR model* [i.e., an AR(1) model]. In general, an AR(p) process is described by

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t. \quad (5)$$

Figure 1C shows an AR(1) series with $\phi_1 = .7$, indicating a positive relationship between successive observations. For an RT experiment, this would indicate that fast responses tend to follow fast responses and slow responses tend to follow slow responses. The effect of a partial dependency (i.e., $\phi_1 = .7$) instead of a complete dependency (i.e., $\phi_1 = 1$, which gives a random walk process) is to keep the values of the observations within certain bounds, preventing them from drifting to either very high or very low values. The simplicity and concep-

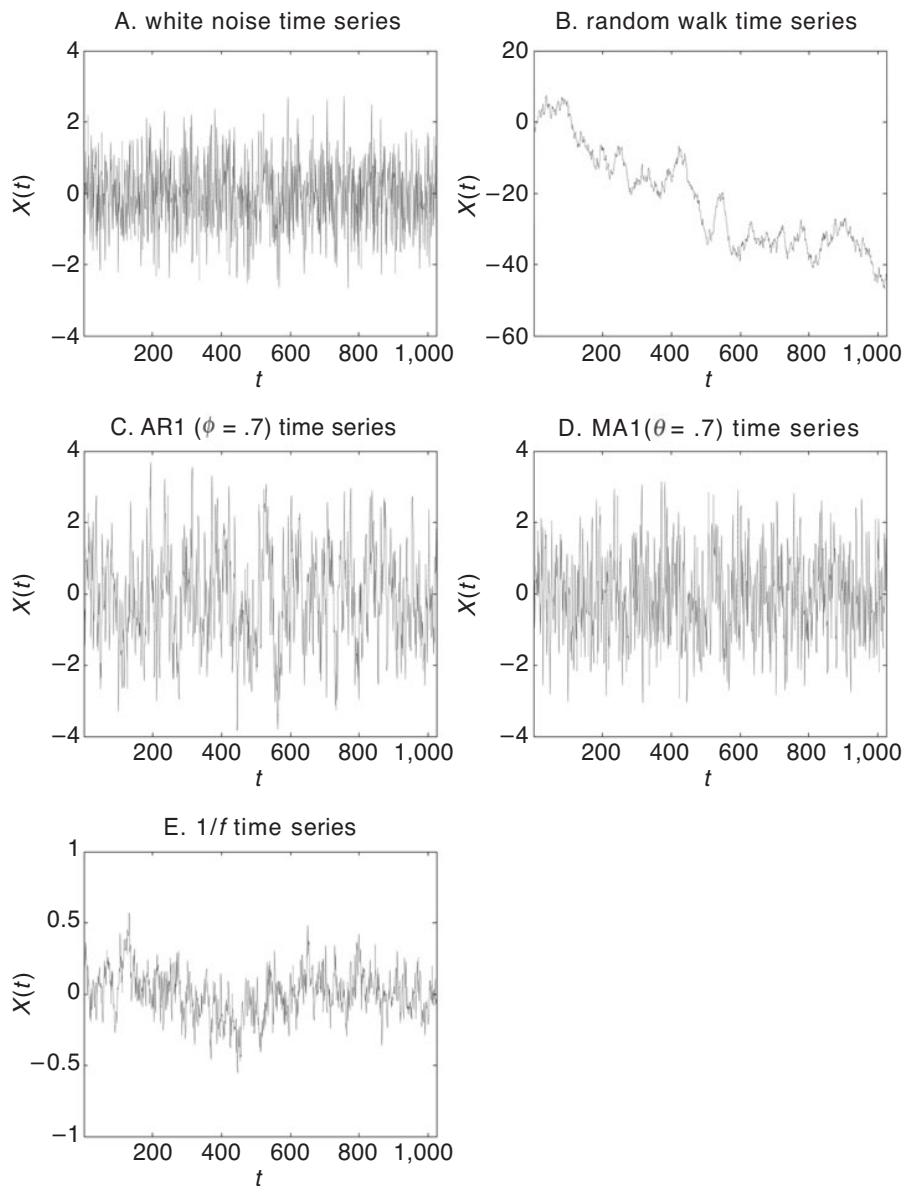


Figure 1. Prototypical time series generated by five models: (A) a white noise process, (B) a random walk process, (C) AR(1) with $\phi_1 = .7$, (D) MA(1) with $\theta_1 = .7$, and (E) $1/f$ noise. See text for details.

tual transparency of the AR(1) model arguably makes it the most often used time series model. We will later use an AR(1) model and additive white noise as the default time series model against which to test more complex models.

The fourth time series (Figure 1D) is a first-order moving average (MA) process denoted MA(1), where

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}. \quad (6)$$

In an MA model of order 1, performance at time t depends on a combination (i.e., weighted average) of the current random innovation ε_t and the previous random innovation ε_{t-1} . The value of previous innovations ε that fall outside of this window comprising the current innovation, and the innovation immediately preceding it are

completely unrelated to X_t . In general, an MA(q) process is described by

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (7)$$

Consequently, for an MA model of order q , the impact of ε_t will be completely absent (i.e., cut off) after q trials. Traditionally, MA processes have been used as filters to remove irregularities and thus to smooth an observed time series (see, e.g., Kettunen & Keltikangas-Järvinen, 2001; Priestley, 1981).

Both the AR model and the MA model consist of linear combinations of uncorrelated variables ε_t (i.e., the purely random process). The difference between the two models is that an AR process absorbs the ε into X_t , meaning that a

value for ε at time t will exert an influence on the values of all future events X_{t+k} , whereas an MA process expresses X_t as a finite combination of previous ε s, such that the random innovation on time t , ε_t , will have no effect on X_{t+k} when k exceeds the order q of the MA model.

Note that a time series need not consist of the observations from only a single type of process. A p th-order AR process and a q th-order MA process can be combined in what is called an ARMA(p, q) process:

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \varepsilon_t + \sum_{r=1}^q \theta_r \varepsilon_{t-r}. \quad (8)$$

The first two terms of Equation 8 correspond to the AR(p) process identified in Equation 5, whereas the last two terms correspond to the MA(q) process of Equation 7 (the middle term ε_t in Equation 8 plays a role in both equations). The ARMA process is the combination of an AR and an MA process, and thus inherits properties of both the AR and the MA components.

As will be explained, the time series generated by the AR and MA processes discussed above are short-range dependent. In contrast, the last time series shown in Figure 1 (panel E) is generated by an LRD process. It is this type of process that is the focus of this article. We will postpone a more detailed quantitative treatment of the identifying features of LRD until after discussing Figures 2–5 (a mathematical description of a statistical model that generates LRD can be found in the Appendix). Some qualitative features of time series with LRD are discussed by Beran (1994, p. 41) and Samorodnitsky and Taqu (1993). First, note that the series in Figure 1E shows relatively long periods of low values and of high values. Mandelbrot coined this the *Joseph effect* in reference to the biblical character Joseph, who was confronted with a long period of plenty followed by a long period of famine (see, e.g., Mandelbrot, 1977; Mandelbrot & Van Ness, 1968). Second, the isolated periods of high values versus those of low values might give the impression that the series is nonstationary (i.e., that its mean and variance change over time), displaying local trends or cyclical behavior. Nonetheless, the overall series looks stationary.

The five time series from Figure 1 differ in the nature of their serial dependencies. The serial structure of time series can be analyzed by correlating the series with a copy of itself shifted over by k observations (cf. Priestley, 1981, pp. 106–110). An autocorrelation function (ACF) can then be constructed by plotting the correlation coefficient between observations at time t and time $t+k$ for increasing values of k . Formally, the ACF is defined as

$$C(k) = \frac{E[\{X_t - \mu\}\{X_{t+k} - \mu\}]}{E[\{X_t - \mu\}^2]} \\ = \frac{\text{cov}\{X_t, X_{t+k}\}}{\sqrt{\text{var}\{X_t\}\text{var}\{X_{t+k}\}}}, \quad k = 0, \pm 1, \pm 2, \dots,$$

where $C(k)$ denotes the value of the autocorrelation at lag k , E denotes expected value, X_t is the value of time

series X at time t , and μ denotes the arithmetic mean of time series X . Figures 2 and 3 both show the resulting ACFs for each type of time series depicted in Figure 1. To more clearly bring out the theoretical properties of the types of time series under consideration and at the same time give the reader a feeling of the noisiness of the ACF, Figure 2 shows the ACF of an individual series, whereas Figure 3 shows the average ACF based on 100 realizations of the same process. Figures 2A and 3A show the ACF of a purely random process: all k correlations between the value of an observation at time t and the value of successive observations $t+k$, $k \neq 0$ are zero. This follows from the fact that a purely random process consists of independent observations.

From an inspection of Figure 1B, it is evident that, in the case of a random walk, the value of an observation at time t can be used to predict the value of the following observation with considerable accuracy. This is reflected in the random walk ACF, shown in Figures 2B and 3B, where the autocorrelations fall off very slowly with increasing lag k .

In contrast, the ACF of the AR(1) process (Figures 2C and 3C) falls off relatively quickly with increasing lag k . Specifically, the ACF of an AR(1) process follows an exponential function:

$$C(k) = \phi_1^{|k|}, \quad (9)$$

where $C(k)$ is the value of the ACF at lag k (i.e., the correlation coefficient between pairs of observations separated by lag k). Because $C(k)$ is a geometric series, the sum of the autocorrelations converges to a finite value (see, e.g., Apostol, 1966, p. 388). More specifically, if $|\phi_1| < 1$, then

$$\sum_{k=0}^{\infty} \phi_1^k = (1 - \phi_1)^{-1}.$$

If $|\phi_1| \geq 1$, the sum of the series diverges. Note that the negative autocorrelations for the longer lags in Figure 2C are due to sampling error: Figure 3C shows that the averaged, less noisy ACF does not have these negative autocorrelations at the longer lags.

The ACF of an MA process (see, e.g., Figures 2D and 3D) of order q is given by

$$C(k) = \begin{cases} \left\{ \sum_{s=k}^q \theta_s \theta_{s-k} \right\} / \left\{ \sum_{s=0}^q \theta_s^2 \right\}, & 0 \leq k \leq q \\ 0, & k > q \end{cases} \quad (10)$$

(Priestley, 1981, p. 137). Note that, as was mentioned above, the autocorrelations are zero when the lag between observations exceeds the size of the averaging window, which is the order of the MA process plus one. This can be seen in the figure as a sharp cutoff at lag 2, where the autocorrelations then remain at zero.

Now consider the ACF of a series that shows LRD. As can be seen from Figures 2E and 3E, the ACF decays more gradually than for an AR(1) process. To be precise, the ACF of a long-range dependent series falls off with lag k as a power function:

$$C(k) = |k|^{-\gamma}, \quad 0 < \gamma < 1. \quad (11)$$

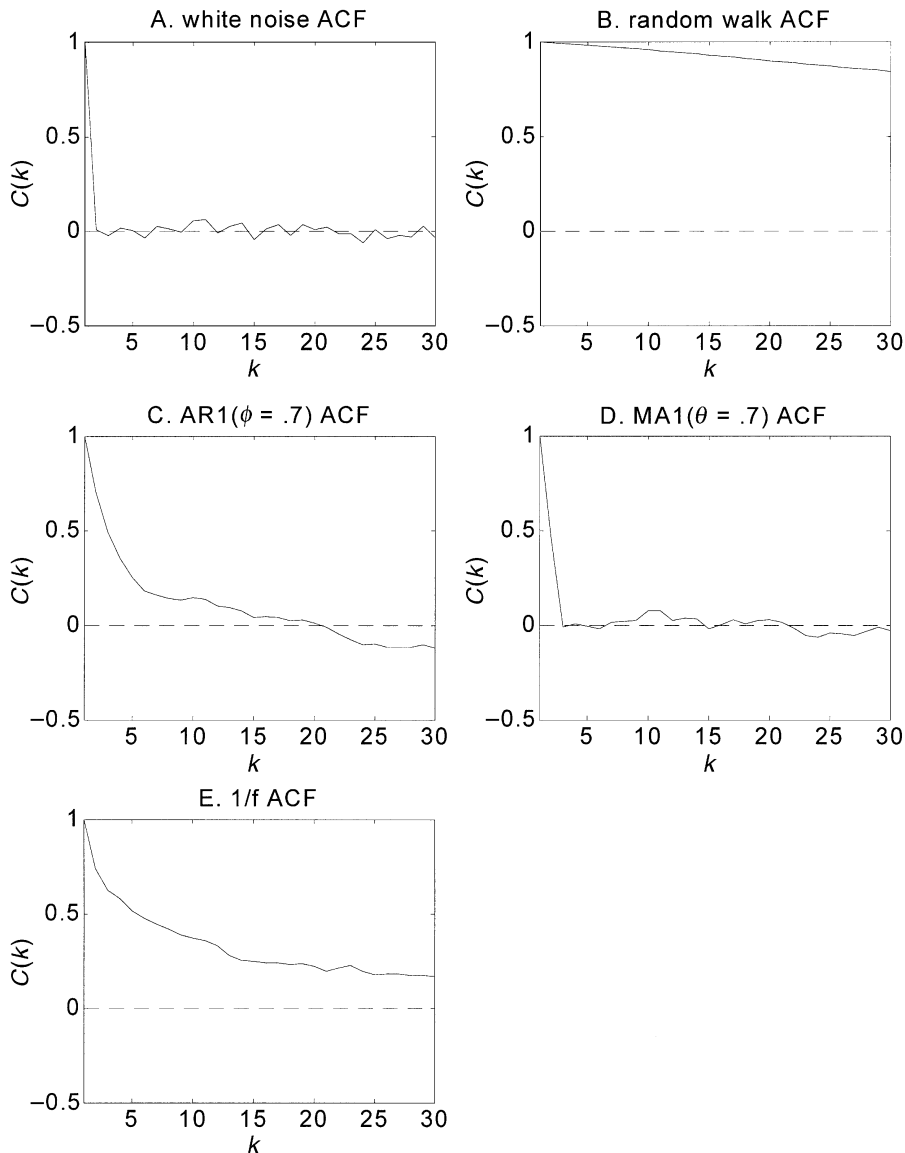


Figure 2. Autocorrelation functions (ACFs) of the time series from Figure 1. For clarity, autocorrelations are plotted for only the first 30 lags. The dashed lines correspond to an autocorrelation of zero.

The autocorrelations of a process adhering to Equation 11 are not absolutely summable (i.e., the sum of the series diverges; see Apostol, 1966, p. 398 for details), so that

$$\sum_{k=1}^{\infty} C(k) = \infty. \tag{12}$$

In the sciences, Equation 12 is one of the widely accepted definitions of a long-range process (see, e.g., Beran, 1994; Samorodnitsky & Taqqu, 1993; but see also Hall, 1997).³ We will mention this definition in an alternative form when we discuss Figures 4 and 5.

So far, we have been concerned with analysis in the time domain. We will now turn to analysis in the frequency do-

main. Any stationary time series (i.e., one whose statistical properties are constant over time; Jenkins & Watts, 1968, p. 147; see also Brockwell & Davis, 1987, pp. 11–14; Kantz & Schreiber, 1997, pp. 13–17) can be reexpressed as an infinite sum of sine and cosine terms, each with a specified frequency and amplitude; this process is called *Fourier analysis* (for an introduction to Fourier analysis, see, e.g., Bloomfield, 2000; Jenkins & Watts, 1968; Priestley, 1981). In physics, a description in the frequency domain is often preferred because it shows the amplitude (or, as is conventional, the squared amplitude or power) for separate frequency bands (a frequency band being a specific interval on the frequency scale). This information is extremely

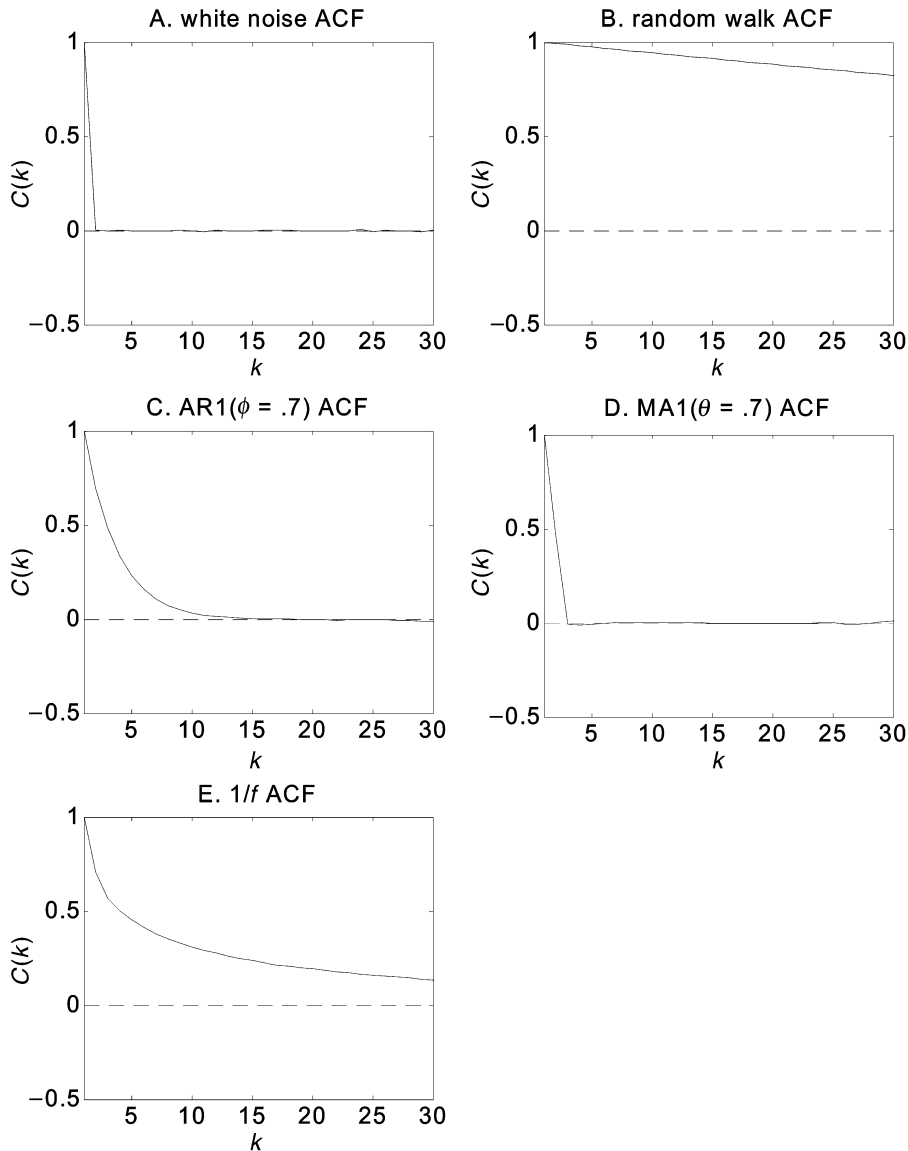


Figure 3. Average autocorrelation functions (ACFs) based on 100 realizations of the processes that generated the time series from Figure 1. The dashed lines correspond to an autocorrelation of zero.

useful for applied purposes; for example, it can be used to construct filters that selectively block out certain frequencies or vibrations, a procedure that is crucial for various practical applications such as radio signal transmission and the construction of airplane wings. In psychology, RT data can also be subjected to analysis in the frequency domain (i.e., spectral analysis; cf. Sheu & Ratcliff, 1995; P. L. Smith, 1990), where frequency is defined as inverse trial number (Gilden, 2001). In this case, high-frequency components correspond to trials that are close together, whereas low-frequency components correspond to trials separated by a large number of observations.

It is a common practice to calculate the power spectrum of a series rather than the standard Fourier transform. The power spectrum function is defined as the squared amplitude of the Fourier transformed times series—that is,

$$S(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} \left| \int_{-T}^T X_t e^{-i2\pi ft} dt \right|^2. \quad (13)$$

For details on the calculation of the power spectrum, see, for instance, Brigham (1974) or Press, Flannery, Teukolsky, and Vetterling (1986).⁴

Following Figures 2 and 3, Figure 4 shows power spectra for the single series of Figure 1, whereas Figure 5

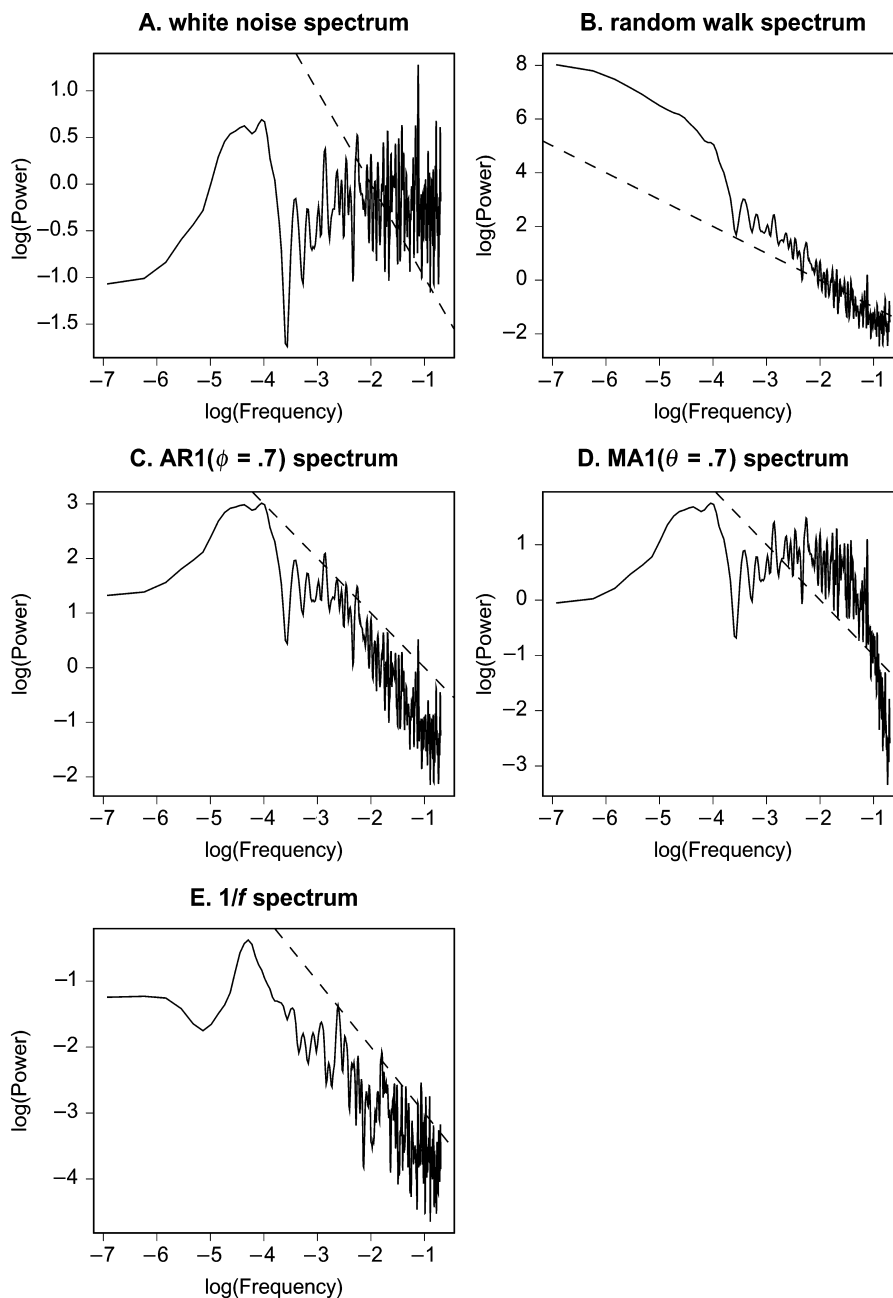


Figure 4. Log-log power spectra of the time series from Figure 1. The dashed lines have a slope of -1 .

shows the average power spectrum based on 100 realizations of the same processes. All spectra reported in this article were smoothed using modified Daniell smoothers of order (3, 5) (see Bloomfield, 2000, for details). The power spectra are plotted on log-log scale; the reason for this is that long-range dependent processes will follow a power function and will thus be linear in log-log coordinates. Figures 4A and 5A show that for a purely random, or white noise, process, the power spectrum is an approximately straight line with a slope of zero. This indi-

icates that each frequency band has the same amplitude, in the same way that white light is composed of light waves of all frequencies (i.e., colors), each with an equal strength.

As can be seen from Figure 1B, the random walk time series is characterized by large “waves”—that is, low-frequency components (those points toward the origin) with high amplitude. Figures 4B and 5B show the corresponding log-log power spectra, which are linear and have negative slopes due to the impact of the low-frequency

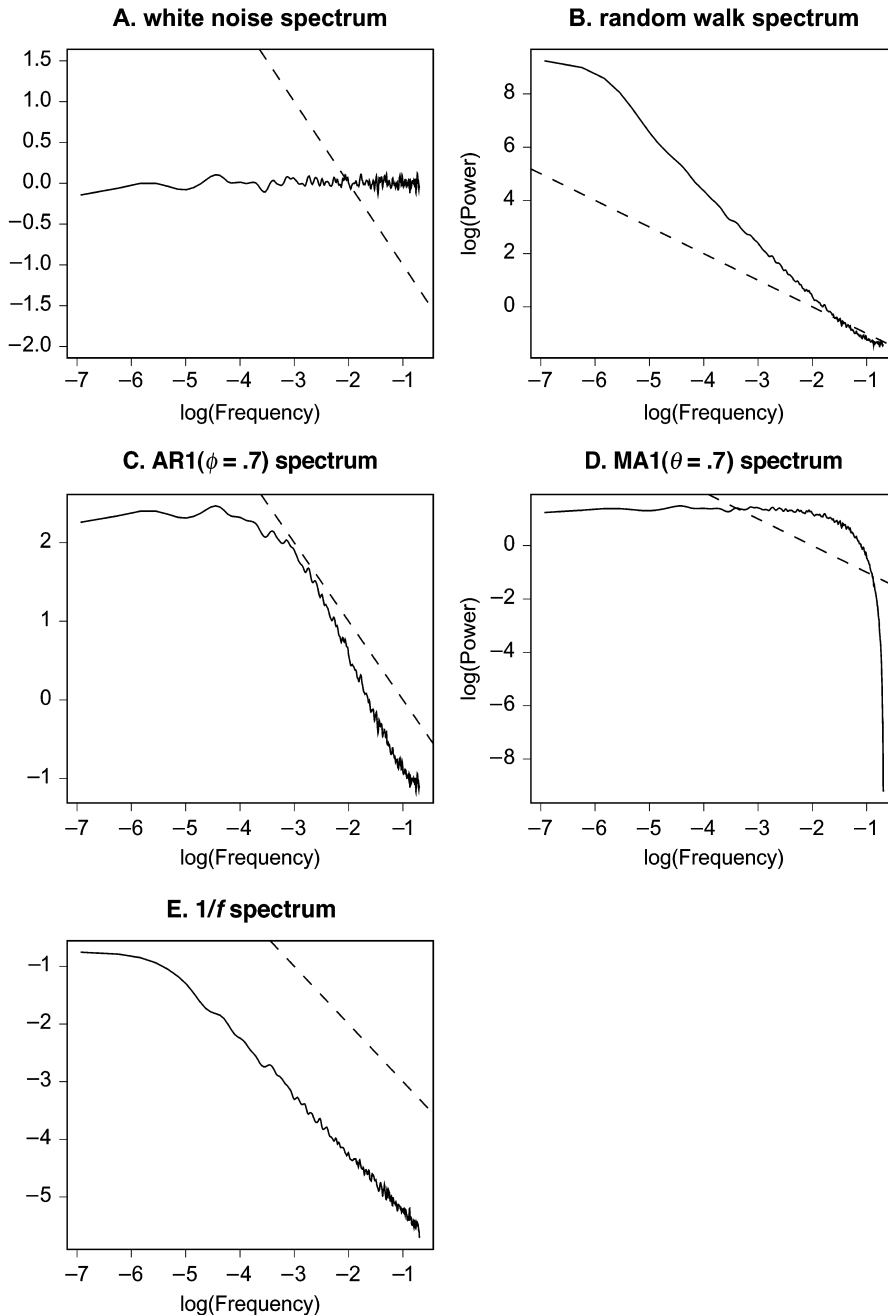


Figure 5. Average log–log power spectra based on 100 realizations of the processes that generated the time series from Figure 1. The dashed lines have a slope of -1 .

components (note that low frequencies are toward the left in the figures). Theoretically, the log–log power spectrum of a random walk is a straight line with a slope of -2 ; that is, $S(f) = 1/f^\alpha$, where $\alpha = 2$ (see, e.g., Kasdin, 1995). Recall that by differencing data generated by the random walk model (with $\alpha = 2$) we can obtain white noise (with $\alpha = 0$), and integrating white noise yields a random walk. Hence, we can see that one-time differencing of a random walk decreases α from 2 to 0, and integrating white noise increases α from 0 to 2. Generally, differencing

increases the power spectrum slope by 2, and integrating decreases the slope by 2 (Pilgram & Kaplan, 1998).

Figures 4C and 5C show the log–log power spectra for the AR(1), $\phi_1 = .7$ process. There are two important points to note about the AR(1), $\phi_1 = .7$ spectrum. One is that more power is present in the low-frequency bands than in the high-frequency bands. Thus, in the AR(1), $\phi_1 = .7$ model, as in the random walk model, SRD (i.e., a relation between the value of an observation on trial t and trial $t-1$) leads to the occurrence of low-frequency

components that have more power (i.e., a higher amplitude) than high-frequency components. However, the second point to note is that the log–log power spectrum for the AR(1) process is not a straight line. Rather, the spectrum curves and flattens out at the low frequencies, reflecting the fact that there is no LRD. Generally, the power spectrum of an AR(1) process crosses over from a slope of -2 at high frequencies (as for a random walk) to follow a white noise power spectrum at low frequencies with a slope of 0 (see, e.g., Hausdorff & Peng, 1996).

The log–log power spectra for the MA(1) process with $\phi_1 = .7$ (Figures 4D and 5D) are very similar to the spectra for the purely random process (Figures 4A and 5A), except that the MA(1) process has a very steep slope at the highest frequencies. This reflects the fact that, for an MA(k) process, the dependencies are zero for lags exceeding the averaging window and are highly dependent for observations within the window. For an MA(1) process, which has a window size of 2, SRD is exhibited only at the very highest frequencies.

Figures 4E and 5E show the log–log power spectra of the time series from the LRD process (Figure 1E). As for the random walk model and the AR(1), $\phi_1 = .7$ model, low-frequency components have more power than high-frequency components. However, the log–log LRD power spectrum can be distinguished from the AR(1) spectrum in that it follows a straight line, here with slope ≈ -1 . The straight-line property, particularly at the low frequencies, mirrors the fact that the ACF of a long-range process is a power function. This means that in the frequency domain a long-range process goes to infinity as the frequency goes to zero: $S(f) \sim f^\alpha$. Time series that exhibit this relation go under various names: *pink noise*, *flicker noise*, *power law noise*, *burst noise*, *low-frequency divergent noise*, and, most commonly, *1/f noise*. The term *1/f noise* implies that $\alpha = 1$, but α is generally taken to vary from 0.5 to 1.5 in the physical sciences literature, and we will therefore use the more general term *1/f $^\alpha$ noise*. Processes that generate *1/f $^\alpha$ noise* by definition generate LRD (the focus of this article), but *1/f $^\alpha$ noise* is also associated with other properties, such as scale invariance and self-similarity, discussed in the next section.

In summary, Figure 3 shows that a power spectrum with low-frequency components that have a higher amplitude than high-frequency components does not necessarily indicate the presence of a long-range process. The AR(1), $\phi_1 = .7$ model has an ACF that falls off exponentially (i.e., that has no LRD), yet the cumulative effect of first-order dependencies results in low-frequency components with high amplitude (see Figures 4C and 5C). In order to distinguish an AR(1) process from a long-range process in the time domain, one has to determine whether the ACF falls off as an exponential function, indicating SRD, or as a power function, indicating LRD. In the frequency domain, a log–log power spectrum that flattens at the low frequencies indicates SRD, whereas a log–log power spectrum that follows a straight line, particularly at the low-frequency end of the scale, indicates LRD.

Defining Characteristics of 1/f Noise

Having given this brief introduction to standard time series analysis, we can now provide a more rigorous definition of a long-range process (cf. Beran, 1994). In the time domain, $\{X_t\}$ is defined as a stationary process with LRD if

$$\lim_{k \rightarrow \infty} C(k) / [m_c k^{-\gamma}] = 1, \quad (14)$$

where $C(k)$ denotes the value of the ACF at lag k , m_c is a constant >0 , and $\gamma \in (0,1)$ gives the (inverse) rate of decay toward zero.⁵ In the frequency domain, $\{X_t\}$ is a stationary process with LRD if

$$\lim_{f \rightarrow 0} S(f) / [m_s f^{-\alpha}] = 1, \quad (15)$$

where $S(f)$ is the power spectral density function, m_s is a constant >0 , and $\alpha \in (0,1)$ is the slope of the line in a log–log plot. From Equations 14 and 15, it can be seen that the defining feature of LRD is not the absolute size of the dependence at short lags, but rather the slow rate of decay of the dependence with increasing lag and, more specifically, the asymptotic behavior of the system as lag tends to infinity. Additional defining features of long-range processes will be discussed below. These sections, up until Estimation of LRD, are not a prerequisite for the material in the remainder of this article and may be skipped on a first reading.

Long-Range Dependence, Scale Invariance, Self-Similarity, and 1/f $^\alpha$ Noise

Up to this point, we have treated LRD and *1/f noise* as more or less synonymous. The focus of our discussion is the interpretation and estimation of long-range correlations in the serial pattern of human behavior. As has been defined above (cf. Equation 15), long-range dependencies show up as a straight line in the log–log power spectrum, $S(f) \sim f^\alpha$ with slope α . Processes with $\alpha = 1$ constitute a special case (i.e., a subset) of processes with LRD. These $\alpha = 1$ processes (i.e., strict *1/f noise*) are scale invariant. As the name suggests, a process that is scale invariant has the same statistical structure across all scales of measurement—that is, the characteristics of the system convey no information about the scale of measurement. Mathematically, this means that for a fixed ratio of frequencies, f_1/f_2 , the integrated power is constant. Thus, if the power spectrum is $S(f) = m/f$, where m is a constant, the integrated power is

$$\begin{aligned} P(f_1, f_2) &= \frac{1}{2\pi} \int_{f_1}^{f_2} S(f) df = \frac{m}{2\pi} \{ \ln(f_2) - \ln(f_1) \} \\ &= \frac{m}{2\pi} \ln \left(\frac{f_2}{f_1} \right). \end{aligned}$$

Scale invariance is not an all-or-none concept, however. For a power function $f(t) = t^\alpha$, the relative change $f(kt)/f(t) = k^\alpha$ is independent of t , and in this sense all power functions lack a characteristic time scale. This means that if a small subset of the series is taken and replotted,

both series would have the same defining characteristics. Popular examples of scale-invariant phenomena are clouds and coastlines, both of which possess the same characteristics at various scales of distance, in such a way that a photo does not reveal whether the picture was taken from 10 m away or from 1,000 m away. Although a weak form of scale invariance will provide the basis of one of the psychologically plausible models for LRD that we will discuss toward the end of this article, scale invariance in the strict sense is not a central issue here. The interested reader should consult G. D. A. Brown, Neath, and Chater (2002) and Maylor, Chater, and Brown (2001) for demonstrations and discussions of scale invariance in human cognition.

Long-range processes are self-similar in the sense that a rescaling of the time axis will leave their distributional properties unaffected (see Figure 1 of Maylor et al., 2001, for an example). A process is self-similar with parameter $H \in (1/2, 1)$ if $c^{-H}X_{ct} = {}_d X_t$, where c is the positive stretching factor, $=_d$ denotes equality in distribution, X_{ct} is the rescaled process, and X_t is the original time series (see, e.g., Beran, 1994; Saupe, 1988, p. 75; Voss, 1988). Note that for a power function such as $f(x) = cx^\alpha$, $f(x)$ is proportional to x^α , whatever the rescaling of x (i.e., multiplication of x by a constant; Schroeder, 1991).

We focus on long-range processes instead of on strict $1/f$ noise because long-range processes are more general and encompass the case of strict $1/f$ noise. Also, the slope of the log-log power spectrum tends to vary, and the label *1/f noise* is deemed applicable for a wide range of α s anyway (from about 0.5 to about 1.5). In the following, we will use the labels *1/f $^\alpha$ noise* and *long-range dependence* (or LRD) interchangeably.

$1/f^\alpha$ Noise and Nonlinearity

A common catch phrase in discussions relating to LRD and scale invariance is that the presence of $1/f^\alpha$ noise is “a signature of dynamic complexity” (Gilden, 2001, p. 33; see also Gottschalk et al., 1995). It is true that $1/f^\alpha$ noise can arise in self-organizing critical systems after perturbation (see, e.g., Gilden, 1997; Jensen, 1998). It has also been claimed that $1/f^\alpha$ noise is due to the accrual of functional fractal ambiguity (Van Orden, Pennington, & Stone, 2001, p. 147), although we find this description unclear. The common interpretation of $1/f^\alpha$ noise as an indication of dynamic complexity might lead one to falsely conclude that $1/f^\alpha$ noise is also an indication of nonlinearity or chaos, since these are two central concepts in dynamical systems theory. We will argue here that the presence of $1/f^\alpha$ noise does not necessarily indicate the presence of nonlinearity or deterministic chaos: In fact, $1/f^\alpha$ noise and nonlinearity are to a large extent orthogonal concepts. This is an important point, not only because the necessary relation between $1/f^\alpha$ noise and nonlinearity sounds (falsely) plausible given the relationship elsewhere between chaotic functions and self-similarity (e.g., the Mandelbrot set; Mandelbrot, 1982), but also because the presence of nonlinear structure in

the dynamics of human cognition has received some attention recently (see, e.g., Busemeyer, Weg, Barkan, Li, & Ma, 2000; Elman, 1998; Heath, 2000; Heath, Kelly, & Longstaff, 2000; Kelly, Heathcote, Heath, & Longstaff, 2001; Kelso et al., 1998; Ploeger, van der Maas, & Hartelman, 2002; Port & van Gelder, 1995; Pressing, 1999a; Roe, Busemeyer, & Townsend, 2001; Thelen & Smith, 1994; Usher & McClelland, 2001; Ward, 2002).

It is well known that biological units such as the brain and the heart function at least in part on the basis of nonlinear components or processes. Examples of such nonlinear dynamics include interactions between sympathetic and parasympathetic innervations, adaptation and saturation of receptors, changes in gain of feedback systems coupled with delays, and so forth (see Kaplan, 1997, and references therein). It might be tempting, on the basis of the wide occurrence of nonlinearity in biological systems, to dismiss a linear description of an observed biological or psychological time series beforehand. However, the presence of nonlinear components does not guarantee that the observed data from the system as a whole in fact contain nonlinear structure. In order to conclude that a time series includes nonlinear structure, it is necessary to demonstrate that a description by means of the simpler, linear time series methods is inadequate.

It is the case that $1/f^\alpha$ noise can be generated by both nonlinear models (see, e.g., Davidsen & Schuster, 2000; Schroeder, 1991) and linear models (see, e.g., Milotti, 1995). Milotti remarks: “I wish also to stress the importance of having $1/f$ noise from linear processes: this agrees well with most experimental observations of $1/f$ noise statistics, and entails a considerable mathematical simplification.” (p. 3099). More specifically, the ARFIMA model (Granger & Joyeux, 1980; Hosking, 1981, 1984), discussed in the next section, can produce LRD and allows a reformulation either as an infinite-order AR process—that is,

$$X_t = \sum_{k=1}^{\infty} \phi_{t-k} X_{t-k} + \varepsilon_t,$$

or as an infinite-order MA process,

$$X_t = \sum_{k=0}^{\infty} \theta_{t-k} \varepsilon_{t-k}$$

(Beran, 1994, p. 65; Hosking, 1981). This implies that the value of X_t depends on the past in a linear fashion. Wold’s decomposition theorem (Wold, 1938; Priestley, 1981, pp. 755–760) states more generally that any stationary purely stochastic process can be expressed as a linear function of an infinite number of past innovations (i.e., random perturbations) ε .⁶ A similar conclusion is reached when one considers the definition of LRD introduced above. As was discussed, the main definition of LRD is that the autocorrelation function of a series follows a power function. Note that since this deals with (lagged) correlations, this measure is by its very definition linear. Thus, the hallmark of $1/f^\alpha$ noise is that the accuracy with which future observations can be linearly

predicted from current ones falls off as a power function. Indeed, a generally accepted method of testing for nonlinearity is to construct a surrogate time series that preserves all the linear structure of the observed time series (i.e., it matches the time series under scrutiny for nonlinear structure with respect to the ACF and power spectrum) but has no nonlinear structure (for details, see Hegger, Kantz, & Schreiber, 1999; Kantz & Schreiber, 1997; Schreiber & Schmitz, 2000). In this case, the surrogate series is the simpler model (the statistical H_0 hypothesis) against which the more complex model is tested.

The dissociation between nonlinear structure and $1/f^\alpha$ noise is not merely theoretical. Gilden (1997, p. 300) used surrogate data series analysis to test for the presence of nonlinear structure, and although Gilden (1997) did report $1/f^\alpha$ noise, nonlinear dynamics were not detected. The reverse pattern was found by Kelly et al. (2001), who reported evidence of nonlinear dynamics without any strong indication of $1/f^\alpha$ noise. This reinforces the orthogonal distinction between nonlinearity and $1/f^\alpha$ noise and also highlights the importance of statistical testing against a simpler alternative hypothesis, a recurring theme throughout this article.

ESTIMATION OF LRD

An LRD process is characterized by a persistent, slowly decaying dependence of current performance on performance in the past. More specifically, the preceding section showed that for an LRD process the slow decay of serial dependence with the number of intervening trials is defined as following a power function. In other words, for an LRD process both the autocorrelation function (time domain) and the power spectrum (frequency domain) follow a power function. When plotted on logarithmic axes, the power spectrum for an LRD process will therefore be linear. Visual inspection of the power spectrum is the method used to detect $1/f^\alpha$ noise in the physical sciences, and it plays a central role in the detection of $1/f^\alpha$ noise in human cognition (see, e.g., Gilden, 2001; but see also Chen et al., 1997, and Pressing & Jolley-Rogers, 1997). LRD is deemed to be present when the best-fitting linear equation for the log-log power spectrum, whether for a particular subject or averaged over several subjects, has a slope between -0.5 and -1.5 . Alternatively, the slope is compared to its theoretical value for LRD (-1) and to its theoretical value under independence (i.e., the simpler alternative hypothesis, which corresponds to a slope of 0; Chen et al., 1997; Van Orden et al., 2003). This comparison process is usually qualitative, and inferential testing is rarely performed on the data.

Although the alternative hypothesis of a white noise process is acceptable at face value, it does involve strong assumptions about the nature of the measures being used. In particular, it assumes that the measures of LRD under consideration (such as the spectrum slope, which is primarily used) are sensitive only to long-range correlations and are not spuriously affected by short-range correla-

tions. If it turns out that these measures are sensitive to the presence of short-range correlations, so that they are shifted from the theoretical value for white noise only in the presence of short-term correlations, then some corrective procedure must be used when hypothesis testing is performed.

A related problem associated with the use of the spectrum slope is that in fitting a line to the power spectrum, linearity and, hence, the presence of LRD, is assumed beforehand. The slope of the best-fitting line serves only to determine the parameter of this assumed underlying process and to quantify the intensity of the assumed LRD. Thus, if a linear fit has the appropriate slope, $1/f^\alpha$ noise will be deemed to be present regardless of whether or not the true shape of the log-log power spectrum is a straight line. Although use of such a scheme means that $1/f^\alpha$ noise will almost always be correctly accepted as $1/f^\alpha$ noise, the obvious consequence is that a short-range process such as that generated by a simple AR model may often be misidentified as LRD, because its nonlinear log-log power spectrum (see Priestley, 1981, pp. 238–239) can give a linear fit with the appropriate slope. In other words, this procedure carries the danger of extreme Type I error rates, although the Type II error rates may be acceptable.

These comments also apply to an alternative $1/f^\alpha$ fitting method employed by Gilden (1997, 2001; cf. Gottschalk et al., 1995). Gilden noted that in many cases psychological time series do not represent perfect examples of $1/f^\alpha$ noise and argued that part of the variability in the residuals is due to white noise. In order to estimate the amount of white noise “contamination” in the residuals, Gilden (1997, 2001) fitted a two-source model to data from several choice experiments. The two-source model is formalized as residual_{*t*} = $1/f_t^\alpha + \beta N(0,1)$, where β determines the relative contribution of a white noise process N with zero mean and unit variance, and α determines the estimated slope of the long-range process. The fits of the two-source model (see, e.g., Gilden, 1997, 2001) are quite good and generally yield estimated values of $\alpha \approx .7$ (i.e., within the typical $1/f^\alpha$ range). However, the two-source model suffers the same drawbacks that apply to the standard slope estimate, because it assumes beforehand that $1/f^\alpha$ noise is present in the time series and, hence, does not test whether or not the data show LRD. That is, an alternative model without LRD might also fit the data well.

A final problem with the spectral slope method (as well as many other LRD estimators) is that its discriminatory power is severely reduced when the time series under scrutiny is contaminated by independent (i.e., additive) white noise. To illustrate this point, Figure 6A plots the log-log power spectrum for a $1/f^\alpha$, $\alpha = 1$ process with independent white noise added to it. As can be seen by comparing the spectral shape to the straight line that has a slope of -1 , adding white noise affects mostly the high-frequency components; the high-frequency end of the spectrum is practically flat, whereas the low-frequency end still shows the straight-line property characteristic of

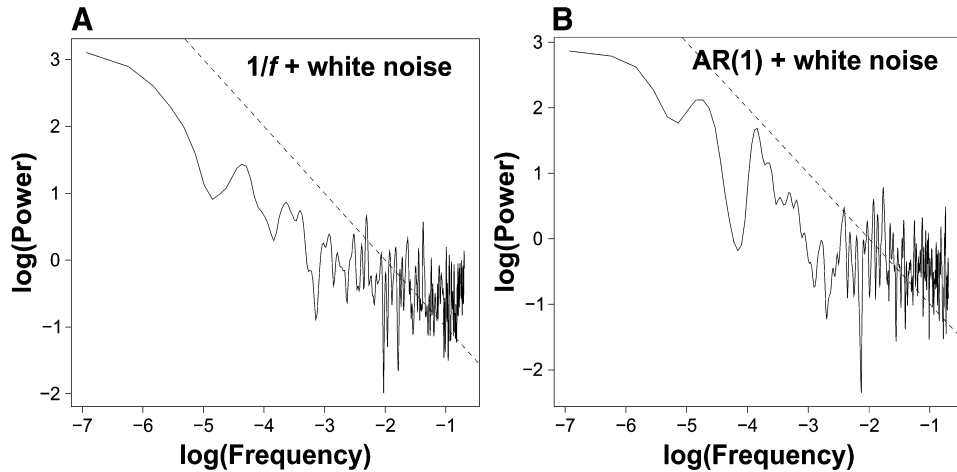


Figure 6. Example log–log power spectra for two different processes in the presence of an independent white noise source. (A) $1/f$ process. (B) AR(1) process plus independent white noise [i.e., ARMA(1,1), $\phi_1 = .85$, $\theta_1 = -.329$]. The dashed line has a slope of -1 .

$1/f^\alpha$ noise. Figure 6B shows the power spectrum of an AR(1) series also contaminated by a white noise source. As we will see later, an AR(1) process plus white noise can be represented as an ARMA(1,1) process, and hence we simulated an AR(1) process plus noise by fitting the parameters of an ARMA(1,1) process to the $1/f^\alpha$, $\alpha = 1$ series, whose spectrum is shown in Figure 6A. From a comparison of the left and right panels, it is clear that, on the basis of the spectral slope method, it is very difficult to distinguish $1/f^\alpha$ noise from AR noise when a white noise source is present. To demonstrate the generality of this result, Figure 7 plots distributions of estimated spectral slopes resulting from a Monte Carlo simulation. In this simulation, 1,000 long-range time series were generated by adding standardized white noise to spectrally synthesized $1/f^\alpha$, $\alpha = 1$ noise (cf. Saupe, 1988). A second group of 1,000 time series was generated by a short-range ARMA(1,1) $\phi_1 = .896$, $\theta_1 = -.691$ process [i.e., AR(1) plus independent white noise]. The values for the ϕ_1 and θ_1 parameters were determined by first fitting an ARMA(1,1) model to each of the $1/f^\alpha$ time series and then averaging the resulting 1,000 ϕ_1 and 1,000 θ_1 parameter values. A third group of 1,000 time series was generated by reducing the SRD produced by the ARMA(1,1) model above—that is, the ϕ_1 and θ_1 were halved to give an ARMA(1,1) model with $\phi_1 = .448$ and $\theta_1 = -.345$. Finally, a fourth group of 1,000 time series was generated by a white noise process. Each time series contained 1,024 observations.

It can be seen from Figure 7 that the discriminatory power between white noise versus $1/f^\alpha$ noise plus an independent white noise source is excellent. However, it is also evident that the spectral slope measure is spuriously elevated by the short-range ARMA(1,1) process. Moreover, by increasing the ARMA(1,1) dependency parameters ϕ_1 and θ_1 , it is possible to shift the slope distribu-

tion across the entire range, indicating that the spectrum slope is highly sensitive to the parameters of this SRD process. To anticipate, we will demonstrate in a later section how the ARMA(1,1) time series used in this simulation can in fact be reliably distinguished from the $1/f^\alpha$ time series.

This simulation serves to illustrate two main points. First, when testing for LRD it is not appropriate to assume an alternative hypothesis of white noise—that is, the state of no serial correlations at all. It is therefore of primary importance to test the LRD hypothesis against the alternative hypothesis of SRD. Throughout this article, we will use as our short-range simple alternative hypothesis the AR(1) plus added white noise process. That is,

$$X_t = Y_t + W_t, \quad (16)$$

where W_t is an independent white noise process and Y_t is an AR(1) process: $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$. Note that Y_t depends on the sequence of innovations ε_t but is independent of W_t . Hence, the two sources of random variation, ε_t and W_t , cannot be simply reduced to a single source of random activation. It has been shown (e.g., Pagano, 1974) that Equation 16 is formally identical to an ARMA(1,1) process (cf. Equation 8)—that is,

$$X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}. \quad (17)$$

Henceforth, we refer to an ARMA(1,1) model rather than to an AR(1) model plus independent white noise. We would like to stress that the choice of the ARMA(1,1) model as a simple alternative hypothesis is not an arbitrary one. ARMA(1,1) is a low-order, parsimonious model and has a psychologically plausible interpretation as an AR(1) process contaminated by independent white noise. The AR(1) model is arguably the most often used time series model, with well-defined properties (see, e.g., Priestley, 1981), and it has a straight-forward interpretation—that is, performance at time t is determined solely by a

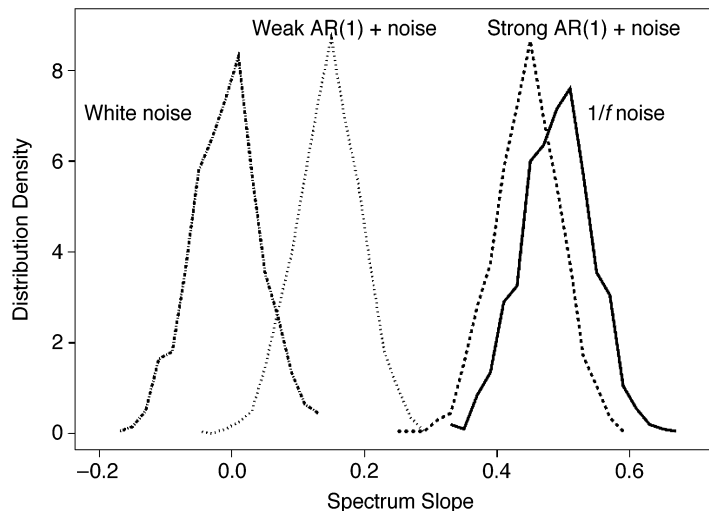


Figure 7. Distributions of estimated spectral slopes for four different processes. From left to right, the processes that generated the distributions are a white noise process; a weakly dependent short-range ARMA(1,1), $\phi_1 = .448$, $\theta_1 = -.345$ process; a strongly dependent short-range ARMA(1,1) $\phi_1 = .896$, $\theta_1 = -.691$ process (where *weak* and *strong* refer to the size of the short-range dependencies); and a long-range $1/f$ process plus standardized white noise. Note that an AR(1) process contaminated by independent white noise can be represented as an ARMA(1,1) process. Each distribution is based on 1,000 time series, each 1,024 observations long.

combination of weighted performance on time $t-1$ and white noise (cf. Equation 4). Because of its conceptual and mathematical simplicity and the ease with which current models of information processing may incorporate variability over time in one or more of their parameters as AR processes, the AR(1) model constitutes a theoretically appealing alternative to a $1/f^\alpha$ process (cf. Laming, 1968).

The second point that is illustrated by the simulation is that the popular spectral slope measure is not satisfactory when the aim is to rigorously test for the presence of LRD. The main reasons for this critical assessment are that the linearity of the spectral shape is assumed beforehand and that the slope method is not robust against the impact of SRD. In the presence of a white noise source, the spectral slope method is unable to distinguish a $1/f^\alpha$ process (i.e., a long-range process) from an AR(1) process (i.e., a short-range process), since either of these processes could give rise to a spectrum with a particular slope for the best-fitting straight line.

Thus, the unsatisfactory performance of the spectral slope method leads us to consider other LRD estimators besides the standard spectral slope method. Over the last 15 years, many different LRD estimators have been introduced and reformalized on the basis of results of various simulation studies. These include the R/S statistic (Hurst, 1951; Mandelbrot & Van Ness, 1968); the adjusted R/S statistic (Lo, 1991; but see Teverovsky, Taqqu, & Willinger, 1999); the spectral slope method using (1) differential weighting of frequencies (Robinson, 1995), (2) low frequencies only (Geweke & Porter-Hudak, 1983), or (3) a

smoothing window that increases in size with increasing frequency (Pilgram & Kaplan, 1998); the variance plot measure (see Beran, 1994); wavelet LRD estimators (see, e.g., Abry & Veitch, 1998); and the Allan variance (see, e.g., Allan, 1996), among many others. A substantial number of papers have been published in which the bias and reliability of several of these LRD measures are examined (e.g., Bassingthwaite & Raymond, 1994; Geweke & Porter-Hudak, 1983; Pilgram & Kaplan, 1998; Schepers, van Beek, & Bassingthwaite, 1992; Taqqu, Teverovsky, & Willinger, 1995; see Baillie, 1996, and Beran, 1994, for overviews). However, most of these studies consider only the bias and efficiency with which a particular LRD measure estimates the intensity of pure long-range processes and therefore assume beforehand the presence of such processes. Thus, these studies generally do not address the issue of hypothesis testing or take into consideration the contaminating effects due to the presence of SRD. Although the theoretical values for a white noise process are known for all measures of LRD, it is rare for the estimated LRD intensity value to be compared to the value of an alternative hypothesis (by simulation or by experimentation). Even when this is done (e.g., via a t test; Chen et al., 1997; Van Orden et al., 2003), the alternative hypothesis that has to be rejected (after which it is concluded that the process has LRD) is assumed to be that of white noise rather than that of a short-range process, as was discussed previously.

In addition to the fact that many LRD estimators do not allow for rigorous inferential testing, many approaches use (arguably arbitrary) cutoffs to selectively consider

only a low-frequency subset of the data. Theoretically, this focus on low-frequency components is well motivated, since the LRD phenomenon is defined in terms of its behavior as frequency approaches zero (cf. Equations 14 and 15). However, the choice of where high-frequency components end and low-frequency components begin is often not straightforward. In addition, discarding most of the data can lead to efficiency losses for the LRD estimators in question (see Baillie, 1996, for similar comments and a more detailed overview).

Given these shortcomings of most current measures of LRD, we were concerned with finding a measure that was relatively robust to the presence of SRD or, alternatively, one that explicitly accounted for SRD in its estimation routine. A method that meets the second of these requirements, ARFIMA time series modeling, is one of the most widely used, flexible, and principled parametric methods for simultaneously estimating both short-range and long-range processes. ARFIMA modeling has a number of desirable features that set it apart from most other LRD estimation methods, but to our knowledge it has not yet been applied in cognitive psychology. We will discuss the ARFIMA modeling framework and its advantages in the next section.

ARFIMA Modeling of LRD

In this section, we provide a general discussion of estimating SRD and LRD using autoregressive fractionally integrated moving-average (ARFIMA) time series modeling. For a more detailed treatment, the interested reader is referred to the Appendix and to other work (e.g., Baillie & King, 1996; Beran, 1994; Granger, 1980; Hosking, 1981, 1984; Kasdin, 1995; Ninness, 1998). The ARFIMA(p, d, q) model operates in discrete time and incorporates both short-range processes (through the AR parameters p and the MA parameters q) and a long-range process (through the fractional differencing parameter d). The ARFIMA(p, d, q) model is a direct extension of the simpler short-range ARMA and ARIMA models, and for ease of exposition we will discuss these simpler models first.

As was mentioned earlier, the ARMA(p, q) process consists of an AR component of order p and an MA component of order q (see Equation 8). The ARMA(p, q) model can be generalized to incorporate the effect of differencing, and this results in an ARIMA(p, d, q) model, of which the added parameter d is an integer that denotes the level of differencing (Box & Jenkins, 1970). For instance, recall that differencing a random walk series (see, e.g., Feller, 1968) results in a white noise series. Hence, the white noise series is given by ARIMA(0, 0, 0) and the random walk series is given by ARIMA(0, 1, 0). Because of its transparency and flexibility, the family of ARIMA(p, d, q) models has been applied often in time series analysis.

Although very well suited for modeling short-range dependencies in time series, ARIMA models cannot give parsimonious descriptions of any long-range dependencies in time series (see Sowell, 1992b, for a discussion).

An ARIMA model would need an infinite number of parameters to accurately describe a long-range process generating a time series of infinite length. However, in recent years the ARIMA model has been generalized to incorporate LRD using only a single parameter. This generalization requires that differencing parameter d be allowed to take on real values rather than the standard integer values (see the Appendix for details). For an intuitive grasp of what the effects of “fractional differencing” are, consider that a white noise process—ARFIMA(0, 0, 0)—has a slope of 0 in the log–log power spectrum, and the random walk process—ARFIMA(0, 1, 0)—has a slope of -2 . Thus, when d is increased from 0 to 1, the spectral slope increases from 0 to -2 . It can be shown that d values between 0 and 1 (i.e., fractional values) result in spectral slopes between 0 and -2 . More specifically, $d = \alpha/2$, where α equals minus the spectral slope. We will restrict ourselves here to $0 < d < 1/2$, for which it can be shown that the ARFIMA(p, d, q) process is stationary and has LRD.

To summarize, the ARFIMA(p, d, q) model has an ARMA(p, q) component to describe short-range processes and a fractional differencing parameter d to quantify the intensity of LRD. If $d = 0$, there is no LRD, and ARFIMA(p, d, q) reduces to an ARMA(p, q) model. If $0 < d < 1/2$, the series is stationary and shows LRD. For $d \geq 1/2$, the series is nonstationary.

In order to test whether the estimated long-range parameter d is significantly greater than 0 for a given time series, the ARFIMA parameters can be estimated using time-domain exact maximum likelihood (EML; see, e.g., Beran, 1994, p. 105; Sowell, 1992a). EML estimation was long believed to be too computationally expensive for practical use, and faster approximate maximum likelihood estimation was used instead (see, e.g., Beran, 1994). However, efficient programming techniques (e.g., the ARFIMA package in the matrix programming language Ox—Doornik, 2001; Doornik & Ooms, 2003) and increased computing resources make EML estimation perfectly feasible for the time series length used throughout this article (i.e., 1,024 observations). The likelihood function that is being maximized is based on the fit of the model to the autocovariance function; for details, the reader is referred to Doornik and Ooms (2003).

Maximum likelihood estimation (see, e.g., Azzalini, 1996; Fisher, 1950; Myung, 2002; Priestley, 1981; Stuart & Ord, 1991) has a number of desirable properties. In general, maximum likelihood estimation is consistent and asymptotically unbiased (see, e.g., Eliason, 1993). For a consistent estimator, the estimated parameter value approaches the true parameter value as the number of observations increases:

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \delta) = 0,$$

where n is the number of observations, θ is the parameter of interest, and δ is arbitrarily small. For an unbiased estimator, the expected value of the estimated parameter value equals the true parameter value: $E(\hat{\theta}) = \theta$. Dahlhaus

(1989) has shown that EML estimation for ARFIMA models is also asymptotically efficient. An efficient estimator has the lowest possible variance among all estimators. In addition, EML yields standard errors for the estimated parameters, which allows for statistical inference. These standard errors are based on the second derivatives of the likelihood function around its maximum (i.e., the slope of the likelihood function is zero at its maximum, the second derivatives indicate how quickly the slope of the likelihood function goes to zero as the maximum is approached, and large second derivatives correspond to small standard errors). This implies that it is possible to fit an ARFIMA(p, d, q) model to a time series using EML and then test whether the estimated LRD parameter d is significantly greater than 0 (or any other value of interest). Unfortunately, this straightforward method of analysis is complicated by several factors. First, misspecification of the short-range component [e.g., fitting an ARFIMA(0, d , 0) to a time series that was in fact generated by an ARMA(1,1) process] can result in incorrect estimates of the LRD parameter d . Second, even if the short-range component is correctly specified, local maxima in the likelihood function can distort the parameter estimates (see, e.g., Myung, 2002). For instance, consider fitting an ARFIMA(1, d , 1) model to a time series generated by an ARMA(1,1) process. When the starting value for d is based on an exploratory method such as the spectral slope, and the likelihood function has a local maximum around this starting value for d , biased estimates for d will result (i.e., $\hat{d} > 0$). With proper precautionary measures—that is, using a number of different starting values—the problem of local maxima can be greatly reduced. However, even when the fit of a correctly specified ARFIMA(p, d, q) model to a given time series results in \hat{d} significantly greater than 0 (calculated on the basis of the second derivatives of the likelihood function around its maximum at \hat{d}) and $\hat{d} > 0$ is a global maximum rather than a local maximum, LRD is not necessarily present: The goodness-of-fit of the short-term ARMA(p, q) component might be only marginally worse than that of the more complex ARFIMA(p, d, q) model.

In order to resolve this crucial issue, we used our simple ARMA(1,1) alternative hypothesis of only SRD as a point of reference for the more complex ARFIMA(1, d , 1). That is, when parameter estimates derived from the ARFIMA(1, d , 1) model indicated that the long-range parameter \hat{d} differed significantly from 0, we used a model selection procedure to quantify whether the higher descriptive accuracy (i.e., the likelihood) of the ARFIMA(1, d , 1) model over the ARMA(1,1) model warrants the inclusion of the extra-long-range parameter d . Thus, following our aims stated above, we used a competitive modeling framework in which the experimental hypothesis of the presence of LRD was actively tested against the simple alternative hypothesis of only SRD (for a similar approach, see Hosking, 1984).

The model selection tool used throughout this article is Akaike's information criterion (AIC; see, e.g., Akaike,

1974; Burnham & Anderson, 2002). The AIC is the most generally recommended method for model selection purposes in the context of time series analyses (see, e.g., Beran, 1994; Priestley, 1981). The AIC was derived as an unbiased estimate of minus twice the expected log likelihood for each candidate model, and it rewards both descriptive accuracy and parsimony (for a detailed discussion on model selection, see Myung, Forster, & Browne, 2000, and Burnham & Anderson, 2002). The lower the AIC score, the more preferable the candidate model. For an ARFIMA(p, d, q) model, the AIC is given by

$$AIC = -2 \ln(L) + 2(p + D + q + 1), \quad (18)$$

where L is the maximum likelihood and D equals 1 when the long-range parameter d is a free parameter, and D equals 0 otherwise. The number 1 added to $p + D + q$ corresponds to a parameter that estimates the error variance—hence, $p + D + q + 1$ is the total number of free parameters. Underspecified models (i.e., those with too few parameters) will be punished by a lack of descriptive accuracy (i.e., low maximum likelihoods), and overspecified models will be punished for having too many parameters. Thus, selecting the candidate model that has the lowest AIC score provides a principled way to quantify the fundamental tradeoff between descriptive accuracy and parsimony.

In this particular case, we compare two *nested* models: the ARMA(1,1) model versus its one-parameter extension, the ARFIMA(1, d , 1) model. When two models are nested and the simple model is correct, twice the difference in their maximized log likelihood values follows a χ^2 distribution with the number of degrees of freedom equal to the difference in the number of free parameters (Wilks, 1938). The likelihood ratio test (LRT) can then be used to reject the hypothesis that the simpler model is correct when twice the observed difference in log likelihood falls in the rejection region of the χ^2 distribution (i.e., usually the most extreme 5%). The choice between using the LRT and using the AIC is to a certain extent a subjective matter. We prefer the AIC because it is a general model selection method and is not based on the framework of null-hypothesis testing (cf. Burnham & Anderson, 2002, pp. 337–339). The AIC is one of the standard model selection tools that can also be used for multiple models that might or might not be nonnested. Furthermore, differences in AIC values can be interpreted as strength of evidence (Wagenmakers & Farrell, 2004) and, most importantly, Monte Carlo simulations reported below show that correct model recovery based on AIC is better than that based on LRT.

In the case of nested models that differ in one parameter, as is the case here, the AIC chooses the more complex ARFIMA(1, d , 1) model when minus two times the log likelihood ratio exceeds 2 (i.e., the AIC penalty for having one additional free parameter). In the null-hypothesis framework of the LRT, this corresponds to a Type I error rate of 15.7% [i.e., a probability of .157 of choosing the ARFIMA(1, d , 1) when the simpler ARMA(1,1) model is

in fact true]. Thus, in this case the AIC is less conservative than the LRT. A Type I error rate of 5% according to the LRT is achieved when two times the likelihood ratio is about 3.84.

We now present a Monte Carlo simulation to demonstrate that the ARFIMA approach to modeling LRD is in practice able to distinguish between a short-range ARMA(1,1) process and an ARFIMA(1, d , 1) process based on the AIC. In this simulation, we used two groups of time series that were also used in the simulations reported in Figure 7. To recapitulate, the first group of 1,000 time series was generated by adding standardized white noise to a $1/f^\alpha$, $\alpha = 1$ process, and a second group of 1,000 time series was generated by the short-range ARMA(1,1) $\phi_1 = .896$, $\theta_1 = -.691$ process. The parameters from the ARMA(1,1) process were determined by fitting the ARMA(1,1) process to the 1,000 $1/f^\alpha$ series and averaging the obtained ARMA parameters. Each time series contained 1,024 observations. On the basis of the AIC, the Type I error rate [i.e., erroneously classifying an ARMA(1,1) process as an ARFIMA(1, d , 1) process] was 7.5%, and the Type II error rate [i.e., erroneously classifying an ARFIMA(1, d , 1) process as an ARMA(1,1) process] was 26.2%. These error rates are more than acceptable in both time series and psychological research, and they provide a necessary demonstration of the adequacy of our procedure for detecting LRD.

The simulation results testify to the discriminatory power of the ARFIMA approach, since it is well-known that for series of finite length the ARMA(1,1) process can be difficult to discriminate from a long-range process (see, e.g., Beran, 1994, p. 145). The ARFIMA approach to modeling LRD has several other important advantages that set it apart from most other approaches. The ARFIMA(p , d , q) model is a natural extension of the popular class of ARIMA models (Box & Jenkins, 1970). ARFIMA models allow for the simultaneous estimation of both short-range processes and a long-range process. When the short-range processes are correctly specified, this greatly increases the robustness of the long-range estimator d against spurious elevation by short-range processes. Furthermore, parameter estimation for ARFIMA models can be done using the asymptotically efficient EML estimator that makes it possible to conduct inferential statistics. Finally, there are fast, user-friendly ARFIMA computer programs and code packages for popular computational statistics programs, to facilitate running of simulations and data analysis (see, e.g., Beran, 1994; Doornik & Ooms, 2003; Ooms & Doornik, 1999).

In this section, we have outlined a method for testing the simple alternative hypothesis of only SRD through the use of ARFIMA time series modeling in combination with AIC model selection. We will apply this method both when analyzing the experimental data reported in the next section and when assessing the adequacy of LRD models discussed later.

WHAT CAN $1/f^\alpha$ NOISE TELL US ABOUT COGNITIVE PROCESSING?

The remainder of this article focuses on what $1/f^\alpha$ noise can tell us about cognitive processing. We will address this issue from both an empirical and a theoretical point of view. We first review current empirical evidence for the presence of $1/f^\alpha$ noise in human cognition. As our review demonstrates, the experimental paradigms and manipulations associated with the occurrence of $1/f^\alpha$ -like noise leave open some important questions and are not as conclusive as might be gathered at first glance. These open questions preclude any firm conclusions about underlying mechanisms. Accordingly, we report an experiment designed to investigate some of the more important unresolved issues, at the same time demonstrating ARFIMA-based inferential testing for LRD. To anticipate the main results, the experiment provides some evidence of the presence of LRD across a range of simple cognitive tasks (i.e., simple RT, choice RT, and temporal estimation). In addition, for several participants the intensity for LRD was most pronounced in temporal estimation.

We then present a detailed theoretical analysis of possible explanations for LRD in human cognition and, more specifically, in our experimental results. We present several psychologically plausible models and show how these can produce long-range dependencies. This demonstration also gives an existence proof that $1/f^\alpha$ noise in human cognition is not uniquely associated with any single process or model, and, hence, one should exert considerable caution when using $1/f^\alpha$ noise to infer underlying theoretical constructs.

Experiments That Lead to $1/f^\alpha$ Noise

Although the study of serial dependence in psychology has a relatively long history (see the introduction for references), little attention has been paid to the possible presence of LRD and $1/f^\alpha$ noise, so that a unified literature on this aspect of cognition does not exist. Nevertheless, as we indicated earlier, exploration of $1/f^\alpha$ noise has recently been undertaken for various tasks and factors.

The most notable of these endeavors has been piloted by Gilden and colleagues (Gilden, 1997, 2001; Gilden et al., 1995; Gilden & Wilson, 1995a). For example, Gilden (2001) presented experiments involving a range of psychological tasks, including mental rotation, lexical decision, rotation search, and translation search. For all these tasks, Gilden observed slopes of the power spectrum between 0 and -1 , giving some indication of the presence of $1/f^\alpha$ noise. In addition, the power spectra for estimation of applied force, angular rotation, and spatial intervals were found to have slopes very close to $\alpha = 1$ (Gilden, 2001, Figure 14). Other researchers have also found serial patterns indicative of $1/f^\alpha$ noise. Van Orden et al. (2003) found slopes between 0 and -1 in pronunciation times and simple reaction times, whereas Chen

et al. (1997) found slopes significantly smaller than 0 in a tapping task.

The experiments conducted by Gilden (2001) and Van Orden et al. (2003) are in many cases standard in cognitive psychology. As was mentioned previously, Gilden (2001) stressed that the amount of variance accounted for by the structure of serial correlations can sometimes exceed the variance accounted for by the experimental manipulation by a factor of two to three. Thus, not only do the above-mentioned results imply serial correlations of a persistent nature (interesting from a theoretical perspective) in situations central to cognitive psychology, but these serial correlations can have a profound impact on task performance, and therefore warrant serious attention.

Although the experiments mentioned above are certainly indicative of the important characteristics of $1/f^\alpha$ noise in human cognition, the research and conclusions based on it are not as clear and consistent as one might gather at first glance. One observation is that empirical investigations of $1/f^\alpha$ noise in human cognition often do not unambiguously show the presence of $1/f^\alpha$ noise. Gilden (2001), for example, reported $1/f^\alpha$ noise in a number of choice RT tasks. Although Gilden (2001) inferred the presence of $1/f^\alpha$ noise in virtually all of his experiments, the majority of experiments reported by Gilden have slopes in the log-log power spectrum that are decidedly less steep than $\alpha \approx 1$. Fitting the reported spectra by hand revealed that for mental rotation (Gilden, 2001, Figure 4), lexical decision, rotation search, translation search (Gilden, 2001, Figure 5), color discrimination, shape discrimination (Gilden, 2001, Figure 8), and discrimination/visual identification judgments with respect to luminance, orientation, the flash of a light, and the absence of a line segment in a square (Gilden, 2001, Figure 16), the slopes of the spectra varied from $\alpha \approx 0.12$ to $\alpha \approx 0.5$. The steeper slopes were obtained when only a subset of the spectrum was fitted, and even with this procedure the mean slope across these particular experiments averages $\alpha \approx 0.3$, in many cases clearly outside the typical $1/f^\alpha$ range of $0.5 < \alpha < 1.5$. It should be noted that a model assuming white noise contamination of a pure $1/f^\alpha$ spectrum (i.e., the two-source model discussed above; Gilden, 1997, 2001) sometimes does yield estimated slopes steeper than -0.5 . However, this method of data analysis is indirect, and in addition it is spuriously effected by SRD (in much the same way as is the standard spectral slope method).

Open questions also exist with respect to findings from different research groups and the conclusions drawn from these findings. It has been suggested by Gilden and colleagues that $1/f^\alpha$ noise is related to the controlled aspects of interval production (Gilden et al., 1995) or the formation of representations (Gilden, 2001). In support, Gilden et al. (1995) showed that $1/f^\alpha$ -like noise was most evident in tasks in which the participants were to repeatedly estimate either spatial intervals or time intervals. In contrast, $1/f^\alpha$ noise was reportedly absent in a simple RT task, in which the participants had to respond as fast as possible

to the onset of a visual stimulus. Gilden et al. pointed out that the magnitude of the response is not intentional in the simple RT task (i.e., participants are not required to produce a specific RT) and argued that the null result (i.e., white noise) obtained in the simple RT task constituted evidence for the assumption that $1/f^\alpha$ noise is selectively "associated with the controlled aspects of interval production" (Gilden et al., 1995, p. 1838). However, in a recent study Van Orden et al. (2003) claimed to find $1/f^\alpha$ noise in a simple RT task and reported a slope of the best-fitting line through the log-log power spectrum of -0.68 . This result is seemingly in contrast to the result predicted and obtained by Gilden et al. It should be noted that in the simple RT experiment by Gilden et al., the participants had to respond manually, whereas the participants in Van Orden et al.'s (2003) experiment had to respond vocally (i.e., by triggering a voice-key). However, it is not clear how this procedural difference could account for the conflicting results. It is also noteworthy that Gilden et al. reported an extremely short mean RT of 100 msec. Previous research has shown that human participants cannot respond faster than 100 msec to the onset of a stimulus, and hence 100 msec has been termed the "irreducible minimum" on RT (see, e.g., Woodworth & Schlosberg, 1954).⁷ It is possible that a substantial proportion of simple RTs in the experiment from Gilden et al. are attributable to anticipations (i.e., initiation of a response before the onset of the stimulus), obscuring the interpretation of the results. Alternatively, D. L. Gilden (personal communication, February, 1, 2004) believed the short mean RT was in fact due to analyses that subtracted a constant estimated motor RT from the data. In addition, note that since the magnitude of the response is not intentional for either simple RT tasks or choice RT tasks, Gilden's (2001) claim of $1/f^\alpha$ noise in choice RT tasks is inconsistent with the earlier claim (Gilden et al., 1995) that $1/f^\alpha$ noise is selectively associated with the controlled aspects of interval production. We will return to this issue later.

In another experiment, Gilden (2001, Figure 11) used a task-switching RT paradigm, using different speeded classification tasks. The results showed that in the mixed condition (i.e., that in which two tasks were switched at random and a cue indicated the current task) the slope of the best-fitting line in the log-log power spectrum was close to zero, revealing no evidence of LRD. In the fixed conditions, however (i.e., where only one task was to be performed throughout the experiment), low-frequency components had more power than high-frequency components. Gilden (2001, p. 43) interpreted this pattern of results by saying, "only when mental set can be consistently maintained are there long-term memory effects over the history of reaction time residuals," relating $1/f^\alpha$ noise to continuity in mental set. Unfortunately, in the task-switching paradigm used by Gilden (2001), consistency in mental set is confounded with task difficulty (i.e., the mixed condition has much higher RTs than the fixed conditions). A more serious problem with the data from Gilden (2001, Figure 11) is that the slopes of the

best-fitting line, which we calculated from the figure by hand, vary from $\alpha \approx .23$ to $\alpha \approx .32$, evidently outside the typical $1/f^\alpha$ range of $0.5 < \alpha < 1.5$ as for the choice tasks. In addition, a visual inspection of the log–log power spectrum leaves the distinct impression that the spectrum levels off at the lower frequencies, which is consistent with a short-range AR process rather than a long-range process. Finally, Gilden and Wilson (1995b) demonstrated that when trials involving the task of dart throwing are interleaved with golf putting trials, performance in the mixed/combined activity is “streaky” (i.e., positive intertrial correlations are observed), whereas performance in the isolated components is not. The latter result is incompatible with the consistency-of-mental-set account of serial dependence in human cognition.

Interpretation difficulties have also arisen regarding the effects of interstimulus interval (ISI). In another set of experiments (i.e., color and shape discrimination), Gilden (2001) manipulated ISI using ISIs of 0.5, 1.0, 1.5, 2.0, 2.5, and 5.0 sec. The results showed that ISI did not affect the intertrial correlations in a speeded object classification task, leading Gilden (2001) to conclude that “... the imposition of delay times has little effect on spectral shape or amplitude” (p. 41). In contrast, Gilden and Wilson (1995a) reported several visual discrimination studies in which a lengthening of ISI reduced the intertrial correlations: “These results indicate that delays of even a few seconds can induce independence in the outcomes between successive trials” (p. 40; for a similar argument, see Gilden & Wilson, 1995b). The latter conclusion is consistent with one of the main points made by Kelly et al. (2001), who argue that a short ISI is an important requirement in bringing about nonlinear dynamics, presumably because a short ISI heightens task demands, requiring participants to exert tighter control over speed–accuracy tradeoffs (see Kelly et al., 2001, for details). Most relevant for the present discussion, Kelly et al. found that a manipulation of ISI caused differences between the slopes in the log–log power spectra, the shortest ISI being associated with the steepest slope. Although the data of Kelly et al. are superficially inconsistent with those of Gilden (2001), we should point out that in addition to manipulating ISI, Kelly et al. (Experiment 2) also varied whether the task was forced-paced or self-paced. The slopes for the forced-paced short ISI, forced-paced long ISI, and self-paced conditions were $\alpha = .31$, $\alpha = .26$, and $\alpha = .16$, respectively. Although the size of the slope differed significantly between the three conditions, the difference in α between the forced choice conditions with a short ISI versus a long ISI was not reliable. Furthermore, Kelly et al. point out that their series may not qualify as $1/f^\alpha$ noise because of the relatively small slopes.

In sum, the current literature has not yet adequately addressed several important questions regarding the experimental tasks and conditions that could lead to the occurrence of $1/f^\alpha$ noise. This state of affairs precludes strong conclusions about the psychological processes in-

involved in $1/f^\alpha$ -like noise. Accordingly, we wished to investigate further the conditions under which LRD is observed. One important and salient factor we wished to address was the effect of type of task on the presence and intensity of LRD. Although tasks such as those involving spatial interval or temporal interval estimation seem to unambiguously show $1/f^\alpha$ noise, it is apparent from the review above that the status of $1/f^\alpha$ noise in the simple RT task is unclear. Choice tasks such as the lexical decision task also seem to provide ambiguous information, in that they generally give α s (i.e., the negative of the spectral slopes) greater than 0 but distinctly less than 1. Therefore, in an attempt to determine the task situations that give rise to LRD, we wished to compare simple, choice, and estimate tasks in a similar experimental format. In particular, to ensure that any differences were not due to differences in stimuli or methodology (which has restricted comparisons between studies in the literature to date), we used exactly the same stimuli and response format in all the tasks and varied only the instructions given to the participants.

Another factor we wished to address was the effect of long or short response–stimulus interval (RSI), which was crossed with task instructions for an examination of its effect on serial correlations. In addition to addressing the inconsistency in results between Gilden (2001) and Gilden and Wilson (1995a), we were interested in this variable because it offers clues as to the processes underlying any LRD. In particular, if the intensity of LRD depends on the temporal contiguity of responses, then LRD should be greater under a shorter RSI.

METHOD

Participants

Six students of Northwestern University participated for a small monetary reward (\$8/h).

Materials, Design, and Procedure

The stimuli were the digits 1, 2, 3, 4, 6, 7, 8, and 9, one of which was presented at the center of a computer screen until a response was registered. A practice phase of 24 stimuli preceded the experimental phase, which contained 1,024 stimuli, each stimulus occurring equally often, the sequence of digits being randomized for each task and for each participant. The experiment featured two within-subjects manipulations: type of task and mean RSI.

The three types of task used were simple RT, choice RT, and temporal estimation. In the simple RT task, the participants were instructed to press the “?” key with the right index finger as soon as they detected the stimulus. To prevent anticipatory responding (see, e.g., Snodgrass, Luce, & Galanter, 1967), we used a variable RSI (see below for details) and presented feedback (“TOO FAST”) for 2 sec following very fast responses (i.e., <100 msec), in addition to verbally instructing the participants to avoid anticipations. In the choice RT task, the participants were instructed to press the “?” key with the right index finger in response to an even number and to press the “z” key with the left index finger in response to an odd number, “as fast as possible without making errors.” The third task was estimation of 1-sec time intervals; the participants were instructed to press the “?” key with the right index finger to indicate when they thought 1 sec had passed since stimulus onset. As in Gilden (2001), no specific instructions were given with respect to

counting (i.e., counting was not discouraged). In all cases, the participants were instructed to keep their fingers on the response key (or, in the case of choice RT, on both response keys) throughout the experiment, to minimize noise due to motor processes. Also, in all cases no feedback was given during the experimental phase for responses that exceeded 100 msec.

The second within-subjects manipulation concerned RSI. Each task (i.e., simple RT, choice RT, and temporal estimation) was performed both with a relatively short RSI and with a longer RSI. The set of short RSIs was randomly drawn from a uniform distribution that extended from 550 to 950 msec, with a mean of 750 msec. The set of long RSIs was obtained by adding a constant 600 msec to the set of short RSIs; hence, the long RSIs varied between 1,150 and 1,550 msec, with a mean of 1,350 msec. The order of the RSIs was randomized for each task and for each participant.

We would like to stress that the only difference between the tasks was in the instruction provided to the participant. The single exception to this rule concerned the practice phase of the temporal estimation task, during which the participants received feedback on their responses (i.e., their estimated time in milliseconds). The design yielded a total of $3 \text{ (task)} \times 2 \text{ (RSI)}$ sessions for the entire experiment, and the order of the tasks was determined using a counter-balanced (Latin square) design. None of the participants performed more than one session a day, and all of them finished the experiment within 1 week.

RESULTS AND DISCUSSION

The results are presented in Table 1 and in Figures 8, 9, and 10. Because the effect of outliers is to reduce the estimated intensity of LRD (e.g., see Beran, 1994, p. 127 for an illustration), prior to analysis we removed outliers in every time series. For all tasks, anticipatory outliers were defined as $RT \leq 100$ msec (the “irreducible minimum”; see Woodworth & Schlosberg, 1954). The definition of slow outliers was task-dependent—that is, $RT \geq 750$ msec in the simple RT task, $RT \geq 1,000$ msec in the choice RT task, and $RT \geq 4,000$ msec in the temporal estimation task. The outlier removal procedure resulted in the elimination of 2.0%, 1.1%, and 0.1% of the data in the simple RT task, the choice RT task, and the temporal estimation task, respectively. Analyses without outlier removal yielded results qualitatively and quantitatively similar to those of analyses with outlier removal. Error responses in the choice RT were not discarded,

since this could potentially disrupt the pattern of sequential correlations.

Figure 8 shows the average ACFs based on the time series for the six conditions. From a visual comparison between the left-hand column (short RSI) and the right-hand column (long RSI), it appears that the pattern of serial dependence is not much affected by the different RSIs used in this experiment (cf. Gilden, 2001). Furthermore, the ACFs for the simple RT task and the choice RT task look quite similar: For both tasks, the lag 1 autocorrelation equals about .2 and decays relatively slowly to low values that are generally still positive up to lag 30. In contrast, the ACFs for the temporal estimation task (Figure 8, bottom panel) have very high lag 1 autocorrelations (i.e., $\approx .5$) and the autocorrelation is about .2 at lag 30. In addition, the visual similarity between the ACFs for the temporal estimation task and the ACF for $1/f$ noise presented in Figure 3E is striking.

The log–log power spectra for the six conditions, averaged across participants, are presented in Figure 9. The results echo those from the ACFs, giving little indication of an effect of RSI (except, perhaps, in the temporal estimation task) and showing similar spectra (slopes $\approx -.3$) for the simple RT task and the choice RT task but different spectra for the temporal estimation task (average slope $\approx -.65$). These observations were confirmed by a 2×3 analysis of variance with RSI and type of task as factors and the size of the slope of the best-fitting line as the dependent variable. There was no effect of RSI [$F(1,5) < 1$] and no interaction between RSI and task [$F(2,10) < 1$]. There was an effect of task [$F(2,10) = 26.03$, $MS_e = .02$, $p < .001$], which t tests confirmed was attributable to the difference between the estimate task and the simple RT task [$t(11) = 6.47$, $p < .001$] and to the difference between the estimate task and the choice task [$t(11) = 6.46$, $p < .001$]. There was no difference in slope between the simple RT task and the choice RT task [$t(11) < 1$].

The analyses presented above give a general impression of the average size and decay of the serial dependence in each of the six conditions. However, the average pattern of serial correlations might not be representative of the pattern for the individual participants, and it has indeed been argued that individual differences in temporal estimation are so great that averaged results are not very representative of any particular participant (Woodrow, 1930, p. 490; Estes, 2002, expressed similar sentiments). The remaining analyses therefore focus on the results from individual subjects.

We used competitive ARFIMA modeling to evaluate statistically the evidence that LRD was or was not present for each individual time series. As was outlined earlier, ARFIMA(p, d, q) time series modeling allows for the simultaneous maximum likelihood estimation of both short-range AR and MA [i.e., ARMA(p, q)] parameters and a long-range parameter d that denotes the level of fractional integration (see the Appendix for details). The d parameter quantifies the intensity of LRD, and $d \in (0, 1/2)$ for stationary LRD processes. The ARMA and

Table 1
Estimated Values for the Long-Range Parameter d for the Individual Time Series From Each Experimental Condition, Provided That the AIC Prefers an ARFIMA(1, d , 1) Model Over an ARMA(1,1) Model

Participant	SS	SL	CS	CL	ES	EL
P1	.13*	.14**	-.17	—	.14	.20*
P2	—	—	—	.14*	—	.45***
P3	.23***	—	—	.12**	—	.36***
P4	.21***	—	—	.23***	—	.45***
P5	—	.04	.11*	—	.40***	.33***
P6	—	.17***	.24**	.29***	—	—

Note—SS, simple response time (RT); SL, simple RT, long RSI; CS, choice RT, short RSI; CL, choice RT, long RSI; ES, temporal estimation, short RSI; EL, temporal estimation, long RSI; P, participant. * $p < .05$. ** $p < .01$. *** $p < .001$. Dashes indicate that the AIC preferred the ARMA(1,1) model over the ARFIMA(1, d , 1) model.

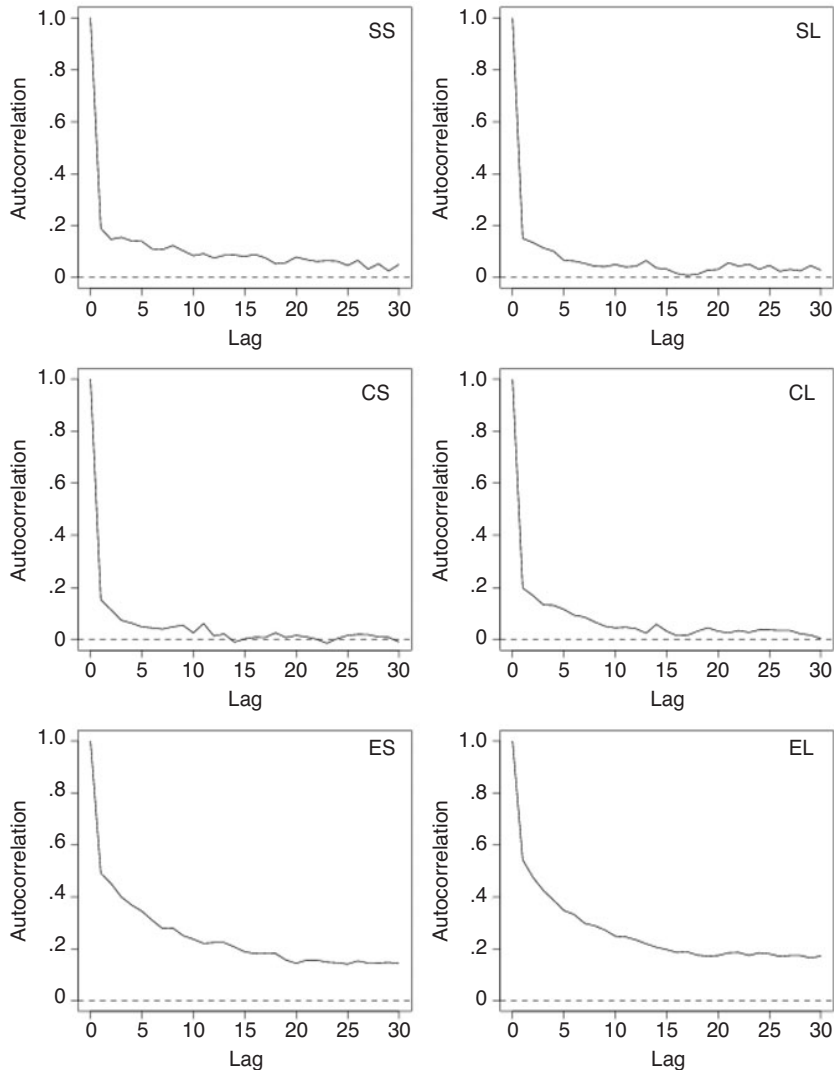


Figure 8. Average autocorrelation functions for the six experimental conditions. SS, simple response time (RT), short response–stimulus interval (RSI); SL, simple RT, long RSI; CS, choice RT, short RSI; CL, choice RT, long RSI; ES, temporal estimation, short RSI; EL, temporal estimation, long RSI.

ARFIMA modeling in this section was performed using the Ox ARFIMA package (Doornik, 2001; Doornik & Ooms, 2003; Ooms & Doornik, 1999) and was checked using the fracdiff package for R.⁸

As in the simulations reported above, the ARMA(1,1) model is our simple alternative hypothesis of SRD: Recall that it can be conceptualized as an AR(1) process plus independent white noise. Following the modeling procedure described in the previous section, we used the AIC (Akaike, 1974) to compare this alternative hypothesis against the hypothesis of the data's being generated by a long-range ARFIMA(1, d , 1) model. As a brief reminder, the AIC quantifies the tradeoff between descriptive accuracy and parsimony and can thus be used to de-

termine whether inclusion of the LRD parameter in the ARFIMA model leads to an increase in log likelihood that is greater than the punishment for adding the extra parameter. Consequently, we fitted both the ARMA(1,1) model and an ARFIMA(1, d , 1) model to each individual time series and chose the model that had the lowest AIC.

Table 1 represents the results of the individual analyses in two ways. First, in cases in which the ARMA gave a lower AIC than did the ARFIMA, the cell is left blank. Second, in cases in which the ARFIMA gave a better fit, the table gives the estimated d value, along with test results (based on the standard error for the parameter estimate) of whether this value differed significantly from zero. Table 1 shows that according to the AIC, the ARFIMA(1, d , 1) is

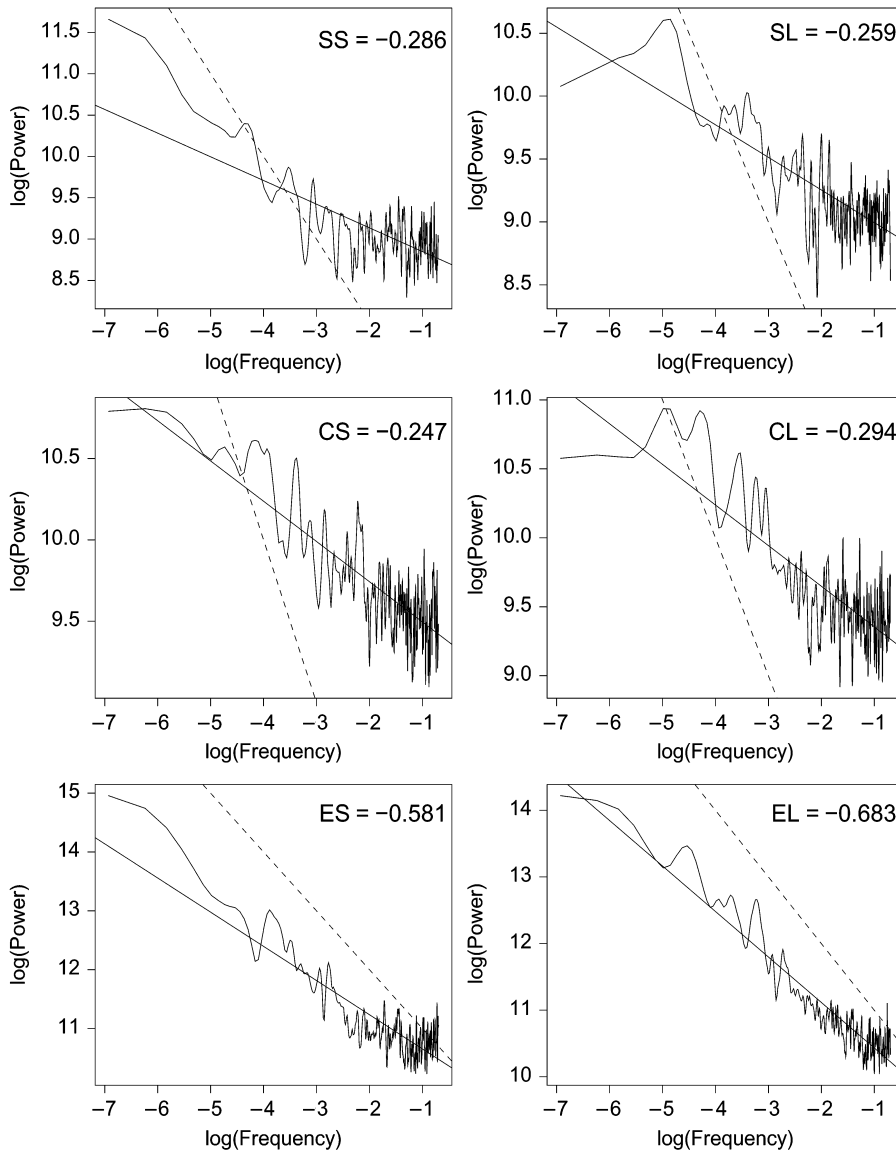


Figure 9. Average log–log power spectra for the six experimental conditions. The dashed line in each plot has a slope of -1 , the solid line is the unweighted least squares best-fitting straight line through the spectrum, and the numbers indicate the slope of the best-fitting straight line. SS, simple response time (RT), short response–stimulus interval (RSI); SL, simple RT, long RSI; CS, choice RT, short RSI; CL, choice RT, long RSI; ES, temporal estimation, short RSI; EL, temporal estimation, long RSI.

to be preferred over the short-range ARMA(1,1) model in 20 of 36 time series. However, in three of the 20 cases in which the ARFIMA(1, d , 1) model was preferred over the ARMA(1,1) model, the resulting d value was not significantly different from zero. Thus, for 17 of 36 series we obtained evidence for LRD. To give an idea of the raw continuous values underlying Table 1, Figure 10 plots the difference in the value of the AIC between the ARFIMA(1, d , 1) and the ARMA(1,1) models, positive differences favoring selection of the ARFIMA(1, d , 1) model. We do not

recommend drawing strong conclusions from AIC differences lower than 2 and consider such values to provide only weak support for the ARFIMA(1, d , 1) model (cf. Burnham & Anderson, 2002).

In both Table 1 and Figure 10, it can be seen that the LRD hypothesis received at least partial support in each of the three tasks (simple RT, choice RT, and temporal estimation). For the simple RT and choice RT tasks, the estimated intensity of LRD is on average much lower than that for the temporal estimation task. Note, how-

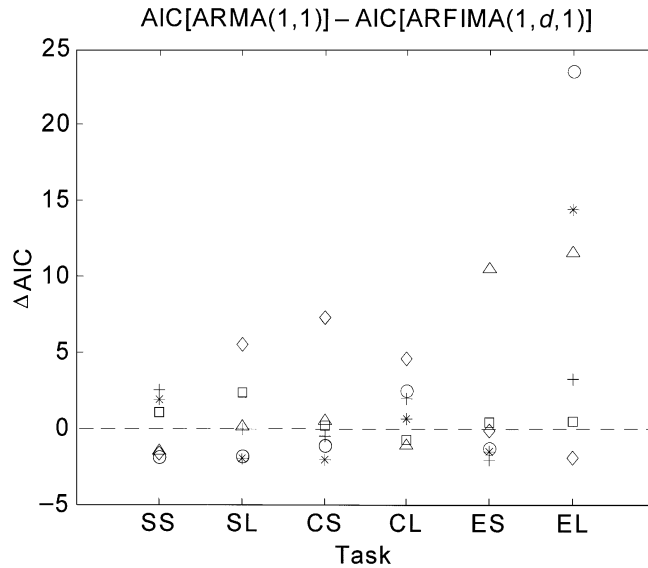


Figure 10. The difference in AIC between the short-range ARMA(1,1) model and the long-range ARFIMA(1, d , 1) model. Positive differences correspond to a preference for the long-range model. Note that if the ARFIMA long-range parameter d does not contribute to the fit at all, the difference in AIC equals -2 . See text for details. Each symbol is associated with a single participant. SS, simple response time (RT), short response–stimulus interval (RSI); SL, simple RT, long RSI; CS, choice RT, short RSI; CL, choice RT, long RSI; ES, temporal estimation, short RSI; EL, temporal estimation, long RSI.

ever, that although \hat{d} is relatively low, significant \hat{d} values are observed for about half the series, and this pattern of results is qualitatively similar for the simple RT task and the choice RT task. The finding that persistent serial dependence can be observed in a simple RT task in which the magnitude of the response is not intentional is at odds with the theoretical position of Gilden et al. (1995). Our results prompted D. Gilden (personal communication, February 1, 2004) to perform a reanalysis of the simple RT data from Gilden et al., and this analysis showed his simple RT spectrum to be in good agreement with the simple RT spectrum obtained in the present study. Thus, both the simple RT data from Gilden et al. and that from the above experiment show some evidence of LRD. These results are consistent with Van Orden et al. (2003); note however, that these serial correlations were smaller than those found in the Van Order et al. (2003) study. It should be noted that when serial correlations are small in absolute size, the more complex model [i.e., the long-range ARFIMA(1, d , 1) model] will generally have difficulty providing a fit to the data that is substantially better than the one provided by the simpler short-range ARMA(1,1) model. In addition, the Monte Carlo simulations reported earlier showed asymmetric Type I and Type II error rates—that is, ARMA models were mistakenly identified as ARFIMA models in only 7.5% of the simulated series (i.e., a Type I error rate of .075), whereas ARFIMA models were mistakenly

identified as ARMA models in 26.2% of the simulated series (i.e., a Type II error rate of .262). Thus, the ARFIMA/AIC method has a conservative bias that favors selection of the ARMA(1,1) model over the ARFIMA(1, d , 1) model. On the basis of these considerations, one might argue that the number of time series classified as ARFIMA series in the simple RT task and the choice RT task (for which the absolute size of the serial correlations was not very high) is actually an underestimate of the true number of ARFIMA series.

Support for the LRD hypothesis is particularly strong in the long RSI condition of the temporal estimation task, in which the long-range ARFIMA(1, d , 1) model outperformed the short-range ARMA(1,1) model for 5 of 6 participants, and the estimated LRD intensity (i.e., \hat{d}) was relatively high. The results of the temporal estimation task are consistent with Gilden (2001), who also found the strongest support for $1/f^\alpha$ noise in estimation tasks. A comparison of the short versus the long RSI condition in the temporal estimation task suggests that the persistence of serial dependencies does not necessarily depend on response events being closely spaced in time.

To summarize, ARFIMA analyses of the experimental data revealed that LRD is present in all three tasks used, although the intensity of the dependence was higher for the estimation task than for the simple RT and choice RT tasks.⁹ The results show that reducing RSI does not in-

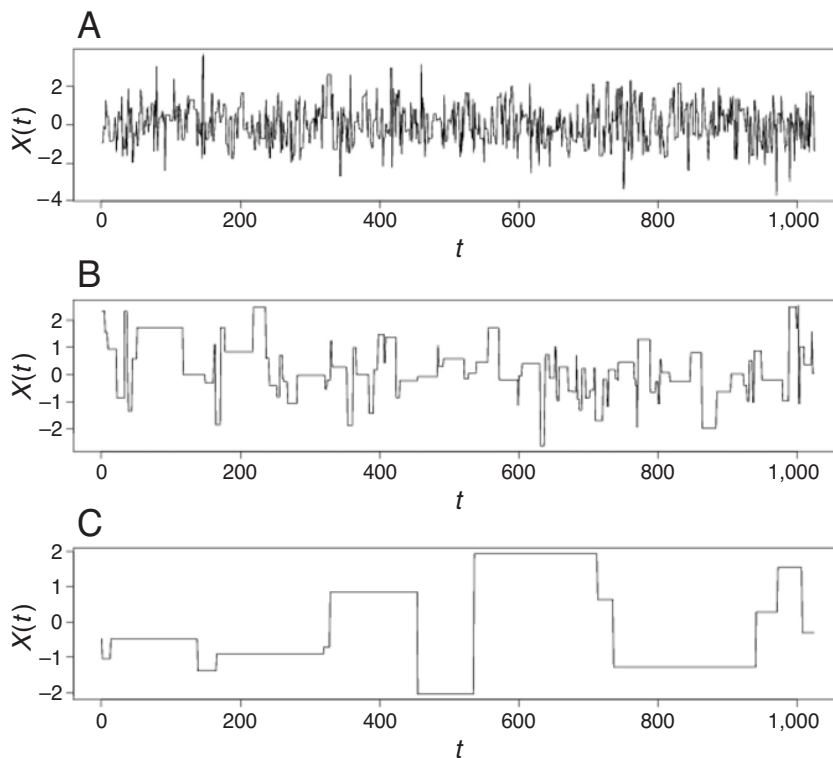


Figure 11. Three time series ($N = 1,024$) that differ in their relaxation rates (i.e., half-lives) τ . At any point in time, the probability of a random change in value (i.e., a switch) is $p_{switch} = 1 - e^{-(1/\tau)}$. (A) $\tau = 1$. (B) $\tau = 10$. (C) $\tau = 100$.

crease the intensity of the LRD; In fact, in the temporal estimation task (and arguably in the choice and simple RT tasks) we obtained evidence to the contrary. Thus, the persistence of serial dependence is not based purely on temporal contiguity. Also, persistent dependence was found in the simple RT task, in contrast to previous claims (i.e., Gilden et al., 1995), but qualitatively consistent with others (Van Orden et al., 2003). In light of the general support for the LRD hypothesis obtained in the experiment, in the next section we discuss how the persistent serial dependencies can be generated and explained by simple principles in psychological models.

Mechanisms for $1/f^\alpha$ Noise

One of our main goals for this article was to study how psychological models can account for LRD in human cognition. We believe such a modeling effort to be crucial to an understanding of the necessary and sufficient conditions for LRD, and even more so because the contribution of the current literature on $1/f^\alpha$ noise in human cognition is arguably more empirical than theoretical.

It should be noted that many models have been proposed to account for LRD or $1/f^\alpha$ noise, especially in the fields of physics and engineering (see, e.g., Handel & Chung, 1993; Marinari, Parisi, Ruelle, & Windey, 1983). However, the models for LRD developed so far are often

not directly applicable to cognitive psychology and have often been developed in terms specific to the domain of application (e.g., magnetic noise and noise in semiconductors, photodetectors, and lasers; for more examples, see Handel & Chung, 1993). The ARFIMA modeling approach outlined earlier is very useful as a statistical tool but is too descriptive to serve as an explanation of underlying psychological processes. Our aim was to present some examples of psychological mechanisms that can generate LRD and to explain how psychological process models can be adjusted to incorporate these mechanisms.

In the process of exploring possible models for LRD, we found that the data for cases in which LRD was deemed to be present (i.e., those in which the ARMA hypothesis gave an insufficient account in comparison with the ARFIMA hypothesis) can be generated by a number of different mechanisms. Thus, even when we accept that a time series may show LRD, its presence alone does not necessarily provide strong constraints for theory development and modeling. To illustrate this important point, we will discuss several general mechanisms that can produce $1/f^\alpha$ noise. Here, we focus on three mechanisms in particular. First, we will show how the summed output of multiple short-range processes with different time scales can yield $1/f^\alpha$ noise within a specific range of frequencies. Second, we will demonstrate that LRD in psychological time series

may arise when participants switch between different strategies during the experiment, particularly in the temporal estimation task. Third, we incorporate the principle of regime switching in a generic sequential sampling model that provides an explanation for LRD in the choice RT task. For all three models, Monte Carlo simulations and ARFIMA analyses support the claim that these models generate convincing LRD, at least for time series 1,024 observations long (which might very well represent an upper limit for the size of an uninterrupted series of observations from human subjects performing typical cognitive tasks). Before outlining the mechanisms and models for LRD in psychological time series in detail, we review two previously suggested models of attentional fluctuation that do not generate $1/f^\alpha$ noise.

Previous Accounts of Serial Dependence in Human Cognition: Fluctuations in Attention

Several researchers have suggested that serial dependence in human cognition is caused by fluctuations in attentional resources (e.g., Busey & Townsend, 2001; Gilden & Wilson, 1995a, 1995b; Laming, 1968; Philpott, 1934). Laming (1968, pp. 117–119) was one of the first to report autocorrelations in RTs (see also Foley & Humphries, 1962). In order to account for data from five experiments showing quickly decaying autocorrelations that remained significantly different from zero up to about lag 6 and that were unrelated to the experimental variables, Laming (1968) proposed that the level of attention paid to the experiment on trial i , A_i , is given by $A_i = A + x_i$, where A is a constant. The deviation from the constant level of attention A , x_i , was assumed to follow a first-order AR scheme (cf. Equation 4): $x_i = \phi_1 x_{i-1} + \varepsilon_i$, where $\phi_1 \in (0, 1)$ and ε is a purely random process. When it is further assumed that the level of attention is inversely related to the observed RT, the ACF is predicted to decay exponentially. Thus, Laming's (1968) model generates SRD, in agreement with his data (see also Botvinick, Braver, Barch, Carter, & Cohen, 2001, pp. 640–643). Of course, this model will not generate the persistent sequential correlations observed in the experiment reported here.

Another model of attentional fluctuation was proposed by Gilden and Wilson to account for streaky performance (i.e., sequential correlations in binary-valued variables such as accuracy) in signal detection tasks (Gilden & Wilson, 1995a) and tasks such as golf putting and dart throwing (Gilden & Wilson, 1995b). According to their attention wave theory, accuracy, or hit rate, varies as

$$\text{hit rate} \propto \sin(2\pi k / L + \theta),$$

where k is the trial number, L is the period, and θ is the phase of the sine wave. This simple model of attentional waxing and waning can explain streaky performance in signal detection tasks, since relatively many hits occur when attention is at a peak and relatively many misses occur when attention is in a trough (similar ideas can be found, e.g., in Philpott, 1934). In addition, Gilden and Wilson (1995a, 1995b) argued that alternative models of

streaky performance based on fatigue and learning are inadequate. Unfortunately, the simple attentional wave model does not generate $1/f^\alpha$ noise, since a sine wave merely corresponds to a spike in the spectrum and does not generate the linear spectral slope that characterizes LRD (this was acknowledged by Gilden, 2001, p. 52). We will later present what can be thought of as a simple model of attentional fluctuation that is able to generate nonstationarity in both hit rate and $1/f^\alpha$ -like noise in temporal estimation tasks (in particular, this model assumes that participants switch between different attentional states).

The preceding examples illustrate that a general verbal explanation of LRD in terms of a waxing and waning of attention, although intuitively appealing, is too superficial: The critical question is not whether attention fluctuations cause serial dependence, but rather the specific manner in which attention must fluctuate to cause the ACF to decay as a power function (the defining characteristic of LRD). In the following, we will discuss several quantitative, psychologically plausible models that demonstrate what specific kinds of fluctuations are sufficient to produce LRD.

Multiscaled Models: Superposition of Short-Range Processes

Perhaps the most straightforward way to produce $1/f^\alpha$ -like noise is to sum short-range processes that have different characteristic time scales (see, e.g., Jensen, 1998, p. 9; Van der Ziel, 1950; note that this is a linear model for $1/f^\alpha$ noise). An intuitively appealing example is given by Richard Voss (in Gardner, 1978): Consider keeping track of the total number of spots on the up faces of three dice. For example, if the up faces of the first, second, and third die show two, five, and three spots, respectively, then the value of the dependent variable equals 10. The first die is thrown anew for every observation, the second die is thrown anew only every now and then, and the third die is thrown anew very rarely. Thus, every die has a different time scale, and it can be shown that the sum of such processes leads to $1/f^\alpha$ -like noise across a range of frequencies, the range being dependent on the number of processes and the parameter of each process (see, e.g., Kasdin, 1995).

To illustrate this process, we now present a simple version of the multiscaled randomness model (Hausdorff & Peng, 1996), developed to provide a general explanation for $1/f^\alpha$ -like noise in biological systems. Hausdorff and Peng argued that in many biological time series overall behavior is influenced by systems operating on widely different time scales. For instance, heart rate fluctuations are regulated beat by beat via the autonomic nervous system, but also show much longer (e.g., circadian) rhythms via hormonal systems (Hausdorff & Peng, 1996). Ward (2002) has presented a similar case for multiscaled randomness in cognitive psychology, identifying component processes as fast fluctuating preconscious processes, slowly fluctuating conscious processes, and unconscious processes that operate on an intermediate time scale.

Pressing (1999b) argued that processes with different time scales are quite common in theories of human cognition. For instance, in memory research one often distinguishes processes associated with fast-acting iconic memory, short-term memory, and slow long-term memory. In research on attention, the presence of automatic, conscious, and sustained attention is often assumed. All of these hypothesized processes typically differ with respect to their time course of decay.

Consider three different systems, S_1 , S_2 , and S_3 , that jointly determine the observed behavioral time series $X(t)$: $X(t) = S_1(t) + S_2(t) + S_3(t)$. In analogy to the die-rolling example discussed above, assume that all three series are switching series—that is, at each unit of time the probability that the system value will change is $p_{switch} = 1 - e^{(-1/\tau)}$, where τ is the time constant of that system. Figure 11 shows the component time series ($N = 1,024$) of the three systems, with $\tau_1 = 1$, $\tau_2 = 10$, and $\tau_3 = 100$. Each of the component series in Figure 11 has an exponentially decaying ACF and hence produces SRD. In fact, a process with relaxation rate τ generates the same serial dependence as a normalized AR(1) process:

$$S_i = \phi_1 S_{i-1} + \sqrt{1 - \phi_1^2} \varepsilon_i, \phi_1 = e^{(-1/\tau)}.$$

The series $X(t)$, obtained by summing the three series from Figure 11, is shown in Figure 12A. Figure 12B shows the average power spectrum based on 1,000 real-

izations of the multiscaled model, and the average ACF is plotted in Figure 12C. Both the slowly decaying ACF and the linearity of the spectrum suggest that the summed series $X(t)$ is long-range dependent for the range of frequencies considered here. On the basis of AIC model selection, the long-range ARFIMA(1, d , 1) model is preferred over the short-range ARMA(1,1) model in 96.7% of the series generated by the multiscaled model.

Methods very similar to the one mentioned here have been suggested to generate $1/f^\alpha$ noise. Granger (1980; see also Beran, 1994) showed that aggregation of k AR(1) processes by random sampling of the AR parameter ϕ_k from a beta distribution can lead to $1/f^\alpha$ noise. This explanation of LRD is popular in the fields of economics and finance (see, e.g., Baillie, 1996). A multiscaled model for LRD in synchronous tapping has also been suggested by Pressing (1999b), in which the outputs from several MA processes operating on the same error process are summed. That is, the current value of an observation is determined partly by MA processes of widely differing window size (Pressing, 1999b). A recent model for the regulation of heart rate through homeostasis (Ivanov, Nunes Amaral, Goldberger, & Stanley, 1998) is also based on the principle that $1/f^\alpha$ noise originates from a combination of processes that operate on different time scales.

The fact that a mixture of exponentials (i.e., short-range processes) can approximate a power function (i.e., a long-

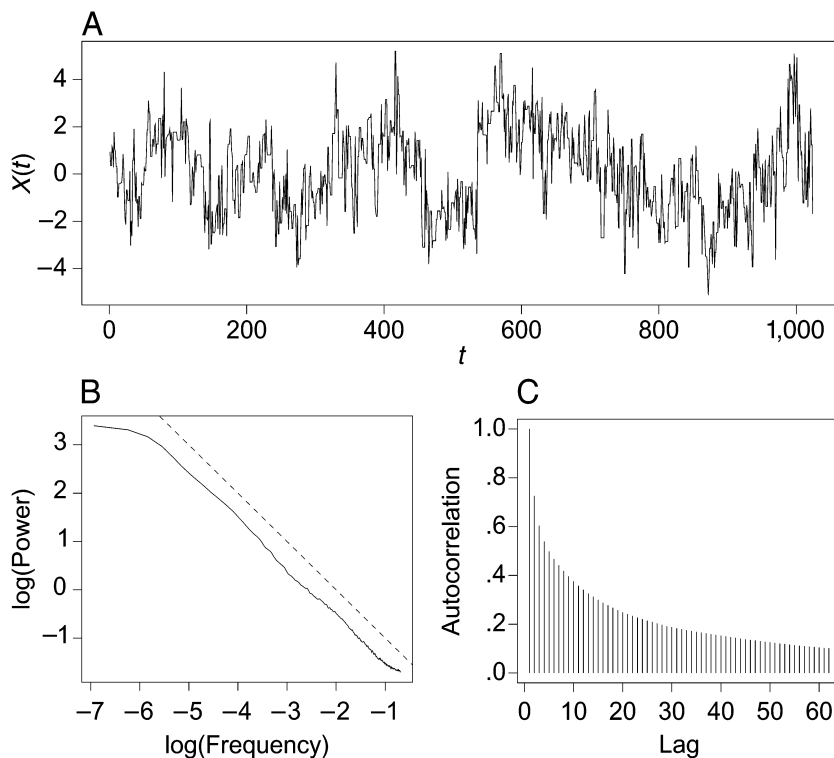


Figure 12. The multiscaled randomness model with $N = 1,024$. (A) An example multiscaled time series obtained by summing the time series from Figure 11. (B) The average multiscaled log-log power spectrum based on 1,000 realizations. The dashed line has a slope of -1 . (C) The average multiscaled autocorrelation function.

range process) is well recognized in psychology. Notable recent work on this issue has focused on the possibility that the power law of practice is in fact an exponential law of practice generating a power function through averaging over participants (Brown & Heathcote, 2003; Heathcote, Brown, & Mewhort, 2000; see also Wixted & Ebbesen, 1997, for a discussion of similar issues with respect to the power law of forgetting). The functions of interest in this paper are just autocorrelation functions rather than standard psychological functions of learning or practice.

The main advantage of explaining LRD using multiscaled models is that the approach is conceptually straightforward and very general. The main disadvantage is that as series length increases, more and more short-range processes with long time scales need to be added to keep the spectrum from flattening, a procedure that is not very parsimonious. For example, extending the length of a time series generated by a multiscaled model with fixed component processes results in fewer detections of $1/f^\alpha$ noise, whereas the reverse pattern would be expected in the case of legitimate LRD. Nevertheless, it can be argued that there is a limit on the length of an undisrupted series that can be gained from an individual, such that only three or four processes will ever be needed to account sufficiently for any time series from cognitive activity in humans.

Shifts in Strategy

Thus far, we have tacitly assumed that the time series under scrutiny for LRD has a constant mean and variance (i.e., second-order or weak stationarity; see Priestley, 1981). The assumption of stationarity is of crucial importance for the measures of LRD discussed in this article, since a violation of stationarity undermines the use of the ACF and the power spectrum. Global or local trends in a time series, such as those due to effects of practice or fatigue, spuriously elevate the ACF (see, e.g., Huitema & McKeane, 1998) and can result in a nonstationary series, with SRD misdiagnosed as long-range dependent (see, e.g., Lobato & Savin, 1998; for details on nonstationarity and its effects on estimating LRD, see, e.g., Beran, 1994, pp. 141–142; Bhattacharya, Gupta, & Waymire, 1983; Dang & Molnar, 1999; Gao, 2001; Giraitis, Kokoszka, & Leipus, 2001; Heyde & Dai, 1996; Künsch, 1986). There is a thin line between nonstationarity and LRD, and it is especially difficult to distinguish the two when the series is relatively short and the nonstationarity takes the form of random drifts over random periods of time (Beran, 1994, p. 142).

Because of the similarities between $1/f$ noise and nonstationarity, it has been suggested that $1/f$ -like fluctuations in a time series might in fact originate from nonstationarity—that is, shifts in the mean or the variance of the observed process. In the literature on economics and finance, such a state of affairs is known as *regime switching* (see, e.g., Kim & Nelson, 1999). For mathematical details on why regime-switching models can mimic LRD, see Diebold and Inoue (2000), Gouriéroux and Jasiak (2001), and A. Smith (2002).

Regime-switching models can offer a psychological explanation for the persistent correlations observed in the temporal estimation task if it is assumed that participants change the strategy they use to complete the task during the course of the experiment. Recall that in the temporal estimation task, the participants had to estimate 1,024 consecutive time intervals of 1 sec, without any breaks. This task is certainly very tedious and may therefore invite the use of different strategies. For instance, Strategy A might be to count (i.e., internally pronounce a number or a word such as “Mississippi”); Strategy B, to visualize a ticking clock; Strategy C, to use the free left hand to tap a rhythm; Strategy D, to construct a sequence of arbitrary phonemes that are judged to take about a second to pronounce internally; and so on. The strategy shift model assumes that every strategy is employed for only a limited period of time (i.e., based on number of trials), given by sampling from a uniform distribution of usage times from 1 to 100; thus, although limited, these periods are quite long on average. Each strategy also has its own temporal criterion a , sampled uniformly from the interval [250, 350], which determines when enough temporal information has accumulated to give a response. This means that the strategy shift model produces local plateaus in performance. It is further assumed that the speed v with which the system approaches the temporal criterion is variable (e.g., using the same Strategy A, participants internally pronounce the same number at different speeds on different trials) and that this variability follows an AR(1) process $v_t = v_1 + \phi_1(v_{t-1} - v_1) + \varepsilon(t)$ with $\phi_1 = .5$, $v_1 = 2$ (i.e., the baseline speed), and $\varepsilon(t)$ white noise with mean 0 and standard deviation 0.1. The RT on trial t is then simply given by a_t/v_t . Figure 13A shows how the temporal threshold shifts across the course of the time series ($N = 1,024$), and Figure 13B shows the speed at which the temporal threshold is approached. The resulting time series $X(t)$ is plotted in Figure 14A. As can be seen from Figures 14B (i.e., the average power spectrum for 1,000 realizations of the strategy shift model) and 14C (i.e., the average ACF), the strategy shift model generates persistent positive serial correlations. Both the approximately linear log–log power spectrum and the power-function ACF suggest that these serial correlations show LRD. This is confirmed using the more constrained AIC model selection, in which the long-range ARFIMA(1, d , 1) model is preferred over the short-range ARMA(1,1) model in 98.4% of the series generated by the strategy shift model.

The basis of the strategy shift model is similar to that of the multiscaled randomness models above, in which several independent processes with different temporal characteristics contribute to the time series. Nonetheless, the strategy shift model of serial dependence in temporal estimation has a number of advantages over the other models presented here. First, it is consistent with recent observations that data from many temporal estimation tasks appear to be nonstationary (Madison, 2001) in that mean performance changes substantially over the course

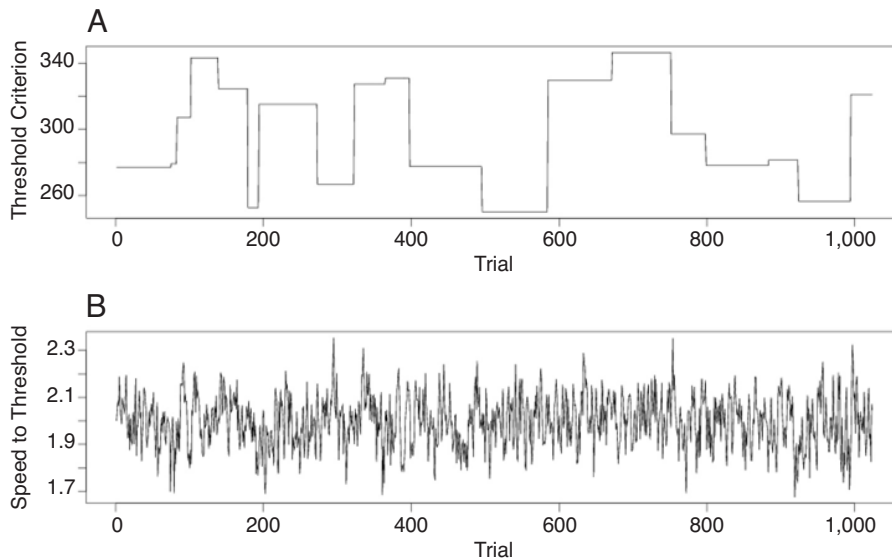


Figure 13. Fluctuations in parameters for the strategy shift model. (A) Example fluctuation of the temporal threshold criterion. (B) Example fluctuation of the speed to the temporal threshold. See text for details.

of the experiment. Second, the model provides an explanation of why the serial correlations are much more pronounced in the temporal estimation tasks than in the simple or choice RT tasks (see also Pressing & Jolley-Rogers, 1997). The estimation task allows participants to actively switch between different strategies, and this leads to persistent serial correlations or LRD. In contrast, the nature of the task is more reactive in the simple and choice RT experiments, meaning that much less of the variance in the task will be due to strategy employment or strategy shifts. A final advantage of the strategy shift model is that it may also be conceptualized as a model for attentional fluctuation, and thus retains theoretical links with earlier explanations of $1/f^\alpha$ noise (Gilden & Wilson, 1995a) and short-range sequential correlations in psychological tasks (see, e.g., Laming, 1968). In this view, attention does not fluctuate in a continuous manner, but rather shifts from one plateau to the other (cf. Van der Maas, Dolan, & Molenaar, 2002, for a similar interpretation in terms of speed–accuracy phase transitions).

Fluctuations in Speed–Accuracy Criteria: A Sequential Sampling Approach

The observation of positive autocorrelations in often-modeled tasks such as choice RT raises the question of how psychological information processing models can be adjusted to account for these data. We will limit our discussion to the class of sequential sampling models, since these models have arguably been most successful in describing behavior in choice RT tasks in detail (see, e.g., Laming, 1968; Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999; Townsend & Ashby, 1983; Van Zandt, Colonius, & Proctor, 2000). The most straightforward

and simple solution to account for serial dependence is to “add on” an attentional effect to the outcome of the decisional process (e.g., as suggested by Laming, 1968). A more principled and interesting approach is to try to incorporate the serial dependence in the model of information processing itself.

A first attempt to account for serial dependence using a sequential sampling model was undertaken by Laming (1979a, 1979b, 1979c). Laming asserted that participants in an RT task have to decide when to begin sampling information from the display. Because of uncertainty associated with the time of stimulus onset, participants may sometimes start the sampling process early and sometimes late. Laming assumed that participants estimate the time of stimulus onset on the basis of a limited number of stored RSIs from previous presentations. The information in the store of RSIs varies slowly over trials, and this mechanism produces an exponentially decaying ACF that provided a good fit to the small and transient autocorrelations observed in several experiments (Laming, 1979a). In order to study how a simple sequential sampling model might generate autocorrelations that resemble those observed in our experiment, we used a random walk model that can be considered a generic sequential sampling model. Note that the random walk of the generic sequential sampling model describes how stimulus information is accumulated within a single trial to give an RT value for that trial, and it should not be confused with the random walk across trials described in the introduction (i.e., Equation 3).

Figure 15 shows a random walk process that provides a description of the decision process on a single trial. At every discrete unit of time, information is sampled from the stimulus and taken as evidence for either Choice A or

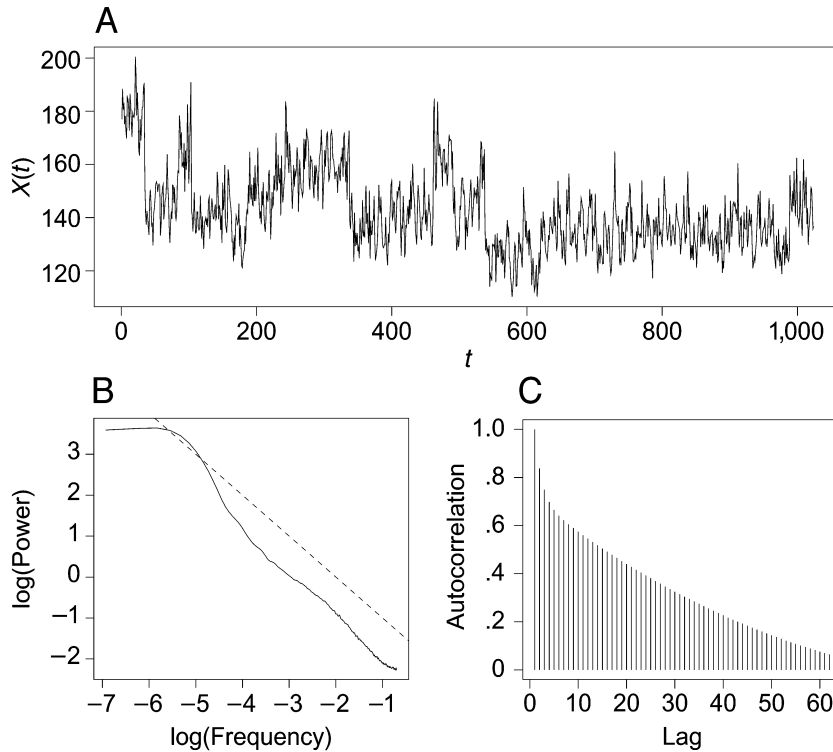


Figure 14. The strategy shift model with $N = 1,024$. (A) An example strategy shift time series based on fluctuations shown in Figure 13. (B) The average strategy shift log–log power spectrum based on 1,000 realizations. The dashed line has a slope of -1 . (C) The average strategy shift autocorrelation function.

Choice B. The sampling process is terminated when one choice alternative has accumulated some criterion amount of evidence more than the other alternative has (the criteria are represented as boundaries on the random walk), and the response that corresponds to the former choice alternative is made. The random walk model for choice RT has a number of parameters that might be assumed to vary over trials. For instance, one might hypothesize that either the probability of correctly evaluating the sampled stimulus information (i.e., p) or the starting point of the process (i.e., z) varies across trials (cf. Ratcliff & Rouder, 1998). By further assuming that the fluctuations in such parameters are not random but follow, say, an AR(1), $0 < \phi_1 < 1$ process, the random walk model can produce positive serial dependence.

In the present implementation, we focused on the parameter a that corresponds to the distance between the boundaries. Boundary separation is often used to model the speed–accuracy tradeoff, in which wide boundaries lead to more accurate responses and longer decision latencies than do narrow boundaries (see, e.g., Ratcliff, 1978; Ratcliff & Rouder, 1998). Thus, the distance between the response boundaries reflects how careful or conservative the system is determined to be. Many factors (e.g., the difference between the perceived and desired levels of accuracy or speed, practice, fatigue, and expected task diffi-

culty) might change the level of conservatism during the course of an experiment. We studied the serial correlations generated by the random walk model when the boundary separation a occasionally shifts (as in the regime-switching model discussed previously), and the drift toward the correct boundary (i.e., the probability of taking a step toward the correct boundary in the discrete random walk model) follows a simple AR(1), $0 < \phi_1 < 1$ process.

Figure 16A shows the result of a realization ($N = 1,024$) of the random walk model where z (i.e., the starting point) is always halfway between the two boundaries and the boundaries have a probability of .01 of switching to a new separation on each trial. When a switch occurred, the new value of the boundary separation was sampled from a uniform distribution of 20–40. Also, p (i.e., the probability of taking a unit step to the top boundary) varied as an AR(1) process (as in the regime-switching model)—that is, $p_t = p_1 + \phi_1(p_{t-1} - p_1) + \varepsilon(t)$, with $\phi_1 = .5$, $p_1 = .65$ (i.e., the baseline p), and $\varepsilon(t)$ white noise with mean 0 and standard deviation .05. To ensure that p did not take on any unreasonable values, it was constrained to lie between .5 and .8; if p drifted outside these values, it was set to the limit which it had exceeded. From Figure 16C, it can be seen that the average ACF of the random walk model drops to less than .2 at the first lag and then decays slowly as the number of in-

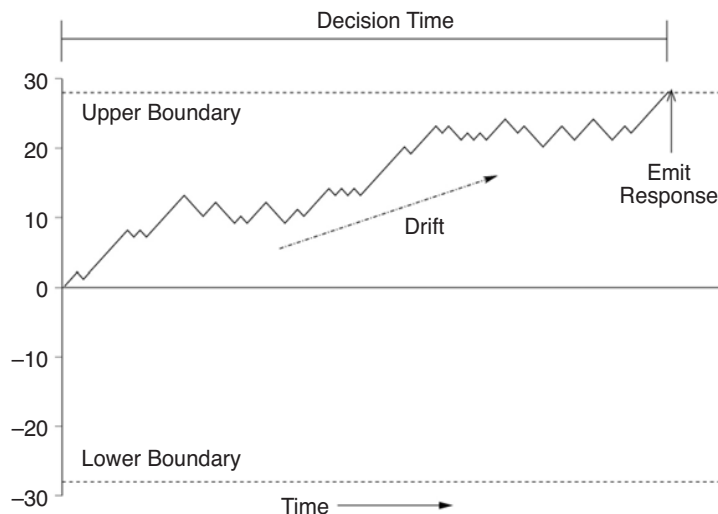


Figure 15. The generic sequential sampling model for a two-choice response time task. See text for details.

tervening trials increases. The absolute value of the ACF at lag 1 is not very large, and this reflects the fact that the dependence caused by the fluctuations in boundary separation a and drift probability p are diluted by the white noise from the stochastic accumulation process. A spectral analysis (Figure 16B) shows that the incline for the lower frequencies is steeper than that for the higher frequencies. On the basis of a visual inspection, the observed spectra for the choice task also appear to have a steeper incline at the lower frequencies than at the higher frequencies (cf. Gilden, 2001, for empirical spectra that are curvilinear with $1/f$ noise for the low frequencies only). However, the main interest here is on the behavior of the sequential sampling model with respect to the lower frequencies: An increase of spectral power at these lower frequencies may indicate that a long-range ARFIMA model is more appropriate than a short-range ARMA model. Indeed, on the basis of AIC model selection, the long-range ARFIMA(1, d , 1) model is preferred over the short-range ARMA(1,1) model in 64.8% of the series generated by the generic sequential sampling model. Thus, by incorporating the principle of regime switching in boundary separation, the generic sequential sampling model yields structure at the lower frequencies that resembles LRD. It is also noteworthy that the average estimated intensity of the LRD (i.e., mean $\hat{d} = .195$) is similar to that empirically observed in the choice RT task.¹⁰

Other Explanations of $1/f^\alpha$ Noise

$1/f^\alpha$ -like noise may be present in many different systems, but it can also be generated by many different mechanisms, a mere subset of which was outlined above. It appears that the presence of LRD per se cannot be used to uniquely identify a specific underlying psychological process or structure. This problem might possibly be overcome in the future, when competing models are developed to such a degree that their adequacy in explaining se-

quential correlations can be qualitatively or quantitatively assessed. Here, we would like to mention two other conceptual approaches that can be used as explanatory frameworks for LRD.

The first type of model for LRD assumes that systems “self-organize” to reach a critical state in which small perturbations can sometimes have dramatic consequences. These models go under the generic label of *self-organized criticality* (SOC; see, e.g., Davidsen & Schuster, 2000; see Jensen, 1998, for an excellent introduction) or *slow-driving interaction dominated thresholded systems* (SDIDTS; Jensen, 1998). The notion of SOC was first introduced to explain $1/f^\alpha$ noise (Bak, Tang, & Wiesenfeld, 1987; but see Jensen, Christensen, & Fogedby, 1989, and Jensen, 1998). Among the phenomena modeled using SOC are earthquakes, evolution, and sandpile dynamics.

Despite the fact that SOC can generate $1/f^\alpha$ noise under certain conditions, Jensen (1998, p. 13) noted, “Although $1/f$ -like spectra might be indicative of critical behavior, they do not guarantee it. There are plenty of ways to produce $1/f$ spectra without any underlying critical state . . .” To our knowledge, the SOC approach has not been applied to $1/f$ noise in human cognition (but see Usher, Stemmler & Olami, 1995, for a population of $1/f$ fluctuating neurons showing SOC on a global scale).

The second type of conceptual framework we mention here is based on the idea that human cognition has become adapted to the statistical structure of the environment. From this point of view, a finding of structured $1/f^\alpha$ noise in human cognition is not surprising, given the supposed ubiquity of $1/f^\alpha$ noise in the physical domain. Anderson and Schooler (1991) showed that both the availability of a memory trace (i.e., performance in a memory task) and the probability of needing that trace (i.e., the statistical structure of the environment) exhibit power law relations with respect to frequency, recency, and the pattern of prior exposures. If the time course of forgetting indeed follows a

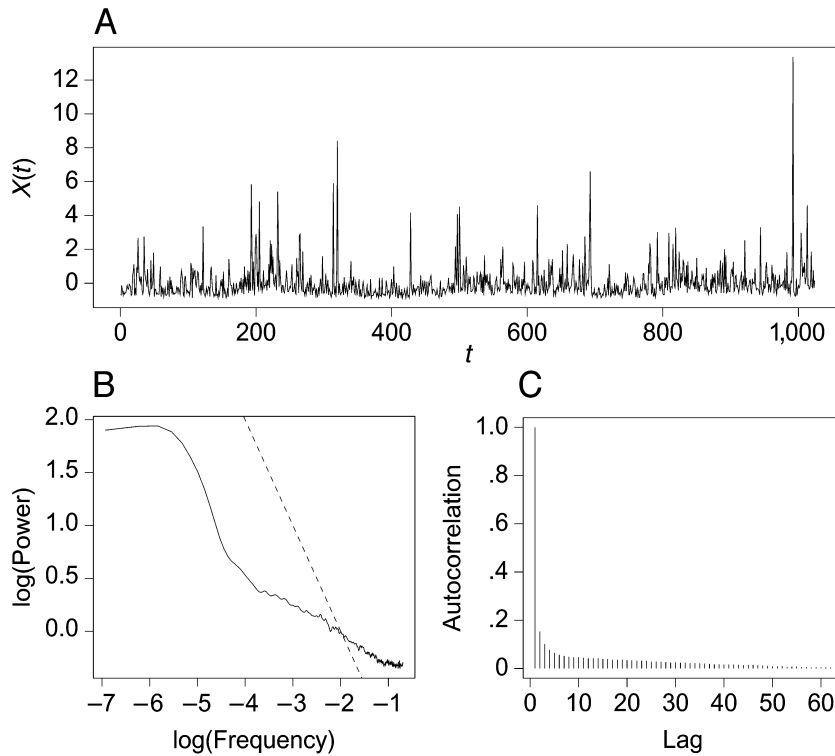


Figure 16. The generic sequential sampling model with $N = 1,024$. (A) An example sequential sampling time series. (B) The average sequential sampling log–log power spectrum based on 1,000 realizations. The dashed line has a slope of -1 . (C) The average sequential sampling autocorrelation function.

power function (see, e.g., Wixted & Ebbesen, 1997), we might speculate that previous performance in estimation tasks can influence later performance in a power law fashion—for instance through feedback based on stored memory traces that contain information on previous performance—thus giving rise to an ACF following a power function (G. D. A. Brown, personal communication, July 19, 2001). Similar ideas regarding scale invariance in human cognition have recently been put forward by G. D. A. Brown et al. (2002).

GENERAL DISCUSSION

The study of LRD has attracted a lot of attention, most predominantly in fields such as economics, physics, and statistics. Recently, LRD, or $1/f^\alpha$ noise, has also been reported across a range of cognitive tasks (see, e.g., Gilden, 2001; Van Orden et al., 2003), and it has been claimed that $1/f^\alpha$ noise provides a new perspective on human cognition. The three major aims of this article were (1) to clarify the defining characteristics of processes that are long-range dependent, (2) to promote the use of SRD as the simple alternative hypothesis in statistical testing for LRD using ARFIMA time series modeling and AIC model selection, and (3) to use psychological models and concepts to explore how LRD might originate in a psychological time series.

With respect to the first objective, we felt that a clarification of long-range processes (or $1/f^\alpha$ noise) was needed in terms of their defining characteristics. Most of the extant literature outside the field of psychology is mathematically dense, published in journals such as *Physical Review* and *Physica*, and hence not easily accessible to the cognitive psychologist. In contrast, most papers published on long-range processes in human cognition tend to present only the most basic result (i.e., a plot of the log–log power spectrum), perhaps under the assumption that the interested reader can consult *Physical Review* for more information. In this article, we have attempted to strike a middle ground between these two extremes, presenting the characteristics of long-range processes without shying away from the most informative details, and referring to other work for rigorous mathematical proofs. The most important result from the extended LRD literature we presented is that the ACF for SRD processes falls off exponentially, whereas the ACF for LRD processes falls off as a power function. In addition, we argued that LRD is not necessarily indicative of nonlinearity or chaos (cf. Kaplan, 1999).

The second objective of this article was methodological. We demonstrated by simulation that the most popular measure for LRD used in psychology (i.e., the assessment of the size of the slope of the best-fitting straight line through the log–log power spectrum) is in

no way conclusive, since it is unduly affected by processes that are short-range dependent. Because spurious elevation in the presence of SRD is a general weakness of most LRD measures, we proposed that in order to establish the presence of $1/f^\alpha$ noise or LRD in human cognition it is necessary to test against the hypothesis of SRD, provided, for instance, by a first-order AR model plus additive white noise [i.e., an ARMA(1,1) process].

We advocated the use of ARFIMA(p, d, q) time series modeling as a principled method that allows for inferential testing of LRD. In an ARFIMA(p, d, q) time series model, the contribution to the observed serial dependencies of both short-range processes (through p AR parameters and q MA parameters) and a long-range process (through fractional parameter d) are estimated simultaneously. Furthermore, by using maximum likelihood estimation of the parameters, it is possible to use the AIC (Akaike, 1974) as a tool for selecting the most appropriate model. That is, the AIC can be used to quantify whether inclusion of the extra-long-range parameter d to a short-range ARMA(1,1) model is warranted on the basis of the amount of increase in descriptive accuracy. We demonstrated by simulation that this procedure is capable of discriminating between noisy long-range and short-range processes that are very difficult to distinguish by eyeballing the power spectra.

The final objective of this article was to elaborate on the conceptual implications of $1/f^\alpha$ noise in human cognition. Empirically, we found many current results on $1/f^\alpha$ noise in the psychological literature to be contradictory, and hence we designed an experiment to investigate two of the most important controversies: the effect of type of task and the effect of mean RSI. The size of the serial dependence was similar for the simple RT and choice RT tasks but was considerably larger for the temporal estimation task. We attribute an earlier reported null result of no serial dependence for the simple RT task (Gilden et al., 1995) to a potential large number of anticipatory responses in that study. Decreasing RSI did not lead to an increase in serial dependence; for the temporal estimation task, decreasing RSI even led to a decrease in serial dependence. From this, we conclude that for the RSIs used here, the serial dependence is not based on temporal contiguity per se. ARFIMA analyses of the experimental data showed support for the presence of LRD in simple RT, choice RT, and temporal estimation. The estimated intensity of LRD was highest for the temporal estimation task.

Theoretically, we demonstrated that given time series of limited length, as are those collected in cognitive psychology, LRD can be generated by a number of different mechanisms. One of the most popular explanations for the phenomenon of LRD in economics and finance is the linear superposition or the aggregation of short-range processes (see, e.g., Granger, 1980). Using ARFIMA analyses and plots of the average ACF and power spectrum, we demonstrated that the summed impact of three short-range processes with different time scales can gen-

erate $1/f^\alpha$ -like noise in a series of 1,024 observations. The constituent short-range processes are not directly observed and, hence, are often left unspecified, but we can speculate that in psychological time series they would correspond to attentional fluctuations, changes in motivation, and the combined effects of fatigue and practice (cf. Ward, 2002).

A second possible explanation for LRD in psychology is that the participants use different strategies throughout the experiment. For the temporal estimation task, we implemented this idea using a regime-switching model and showed that the average ACF and power spectrum, as well as ARFIMA analyses, all indicated that this model produces LRD. The notion that participants use different strategies during the course of an experiment gains credibility when the task either is very tedious or can be performed in many different ways. The temporal estimation task, particularly in the long RSI condition, meets both requirements. The regime-switching model for the temporal estimation task illustrates that LRD might come about through inconsistency of mental set, rather than through consistency of mental set, as suggested by Gilden (2001).

In order to model the small but persistent serial correlations observed in the choice RT task, we used a generic sequential sampling model and incorporated dependencies in the model's parameters across trials that resemble a regime-switching mechanism. More specifically, we assumed that the level of conservatism (i.e., the position on the speed-accuracy tradeoff function) fluctuates over time. When conservativeness, as quantified by the boundary separation, occasionally jumps to a new level, the generic sequential sampling model can generate LRD. In addition, the noise in the accumulation process of the sequential sampling model causes the intensity of the LRD to be lower than that of the regime-switching model for the temporal estimation task, in consistency with the empirical results.

In sum, it is certainly the case that the novel research programs used in the study of long-range processes in human cognition have recently generated a lot of interest. In this article, ARFIMA time series modeling and AIC model selection are shown to provide evidence of the presence of LRD across a range of cognitive tasks (i.e., simple RT, choice RT, and temporal estimation). We have outlined two very general mechanisms for producing LRD (i.e., multiscaled randomness and regime switching) and implemented the notion of regime switching in a generic sequential sampling model to account for persistent correlations in a choice RT task. Thus, this work provides a methodology to diagnose and quantify the presence of LRD in psychological time series and, in addition, our results point to quite general explanations of the presence of LRD across a range of psychological tasks. We believe the use of a robust estimation method and the construction of quantitative psychologically interpretable models can greatly enhance our knowledge of how current performance depends on performance in the past.

REFERENCES

- ABRY, P., & VEITCH, D. (1998). Wavelet analysis of long range dependent traffic. *IEEE Transactions on Information Theory*, **44**, 2-15.
- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.
- ALLAN, D. W. (1996). Statistics of atomic frequency standards. *Proceedings of the IEEE*, **54**, 221-230.
- ANDERSON, J. R., & SCHOOLER, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, **2**, 396-408.
- APOSTOL, T. M. (1966). *Calculus*. New York: Wiley.
- AZZALINI, A. (1996). *Statistical inference*. London: Chapman & Hall.
- BAILLIE, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5-59.
- BAILLIE, R. T., & KING, M. L. (Eds.) (1996). Fractional differencing and long memory processes [Special issue]. *Journal of Econometrics*, **73**(1).
- BAK, P., TANG, C., & WIESENFIELD, K. (1987). Self-organized criticality: An explanation of $1/f$ noise. *Physical Review Letters*, **59**, 381-384.
- BASSINGTHWAIGHTE, J. B., & RAYMOND, G. M. (1994). Evaluating rescaled range analysis for time series. *Annals of Biomedical Engineering*, **22**, 432-444.
- BERAN, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science*, **7**, 404-427.
- BERAN, J. (1994). *Statistics for long-memory processes*. New York: Chapman & Hall.
- BHATTACHARYA, R. N., GUPTA, V. K., & WAYMIRE, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, **20**, 649-662.
- BLOOMFIELD, P. (2000). *Fourier analysis of time series*. New York: Wiley.
- BOTVINICK, M. M., BRAVER, T. S., BARCH, D. M., CARTER, C. S., & COHEN, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, **108**, 624-652.
- BOX, G. E. P., & JENKINS, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden Day.
- BRIGHAM, E. O. (1974). *The fast Fourier transform*. NJ: Prentice Hall.
- BROCKWELL, P. J., & DAVIS, R. A. (1987). *Time series: Theory and methods*. New York: Springer-Verlag.
- BROWN, G. D. A., NEATH, J., & CHATER, N. (2002). *SIMPLE: A local distinctiveness model of scale-invariant memory and perceptual identification*. Manuscript submitted for publication.
- BROWN, S., & HEATHCOTE, A. (2003). Averaging learning curves across and within participants. *Behavior Research Methods, Instruments, & Computers*, **35**, 11-21.
- BURNHAM, K. P., & ANDERSON, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.
- BUSEMEYER, J. R., WEG, E., BARKAN, R., LI, X., & MA, Z. (2000). Dynamic and consequential consistency of choices between paths of decision trees. *Journal of Experimental Psychology: General*, **129**, 530-545.
- BUSEY, T. A., & TOWNSEND, J. T. (2001). Independent sampling vs. inter-item dependencies in whole report processing: Contributions of processing architecture and variable attention. *Journal of Mathematical Psychology*, **45**, 283-323.
- CHEN, Y., DING, M., & KELSO, J. A. S. (1997). Long memory processes ($1/f^\alpha$ type) in human coordination. *Physical Review Letters*, **79**, 4501-4504.
- DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, **17**, 1749-1766.
- DANG, T. D., & MOLNAR, S. (1999). On the effects of nonstationarity in long-range dependence tests. *Periodica Polytechnica: Electrical Engineering*, **43**, 227-250.
- DAVIDSEN, J., & SCHUSTER, H. G. (2000). $1/f^\alpha$ noise from self-organized critical models with uniform driving. *Physical Review E*, **62**, 6111-6115.
- DIEBOLD, F. X., & INOUE, A. (2000). *Long memory and regime switching*. NBER Technical Working Paper No. 264.
- DOORNIK, J. A. (2001). *Ox: An object-oriented matrix language*. London: Timberlake Consultants Press.
- DOORNIK, J. A., & OOMS, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis*, **42**, 333-348.
- ELIASON, S. R. (1993). *Maximum likelihood estimation: Logic and practice*. Newbury Park, CA: Sage.
- ELMAN, J. L. (1998). Connectionism, artificial life, and dynamical systems: New approaches to old questions. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science*. Oxford: Basil Blackwood.
- ESTES, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, **9**, 3-25.
- FELLER, W. (1968). *An introduction to probability theory and its applications*. New York: Wiley.
- FISHER, R. A. (1950). *Contributions to mathematical statistics*. New York: Wiley.
- FOLEY, P. J., & HUMPHRIES, M. (1962). Blocking in serial simple reaction tasks. *Canadian Journal of Psychology*, **16**, 128-137.
- GAO, J. B. (2001). Detecting nonstationarity and state transitions in a time series. *Physical Review E*, **63**, 066202-1-8.
- GARDNER, M. (1978). Mathematical games: White and brown music, fractal curves and one-over-f fluctuations. *Scientific American*, **238**, 16-32.
- GEWEKE, J., & PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221-238.
- 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.
- GILDEN, D. L., & WILSON, S. G. (1995a). On the nature of streaks in signal detection. *Cognitive Psychology*, **28**, 17-64.
- GILDEN, D. L., & WILSON, S. G. (1995b). Streaks in skilled performance. *Psychonomic Bulletin & Review*, **2**, 260-265.
- GIRAITIS, L., KOKOSZKA, P., & LEIPUS, R. (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, **38**, 1033-1054.
- GOTTSCHALK, A., BAUER, M. S., & WHYBROW, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry*, **52**, 947-959.
- GOURIEROUX, C., & JASIAK, J. (2001). Memory and infrequent breaks. *Economics Letters*, **70**, 29-41.
- GRANGER, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14**, 227-238.
- GRANGER, C. W. J., & JOYEUX, R. (1980). An introduction to long-memory models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15-29.
- HALL, P. (1997). Defining and measuring long-range dependence. In C. D. Cutler & D. T. Kaplan (Eds.), *Nonlinear dynamics and time series* (pp. 153-160). Providence, RI: Fields Institute Communications.
- HANDEL, P. H., & CHUNG, A. L. (Eds.) (1993). *Noise in physical systems and 1/f fluctuations*. New York: American Institute of Physics.
- HAUSDORFF, J. M., & PENG, C.-K. (1996). Multiscaled randomness: A possible source of $1/f$ noise in biology. *Physical Review E*, **54**, 2154-2157.
- HEATH, R. A. (2000). *Nonlinear dynamics: Techniques and applications in psychology*. London: Erlbaum.
- HEATH, R. A., KELLY, A., & LONGSTAFF, M. (2000). Detecting nonlinearity in psychological data: Techniques and applications. *Behavior Research Methods, Instruments, & Computers*, **32**, 280-289.
- HEATHCOTE, A., BROWN, S., & MEWHORT, D. J. K. (2000). The power law revealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, **7**, 185-207.
- HEGGER, R., KANTZ, H., & SCHREIBER, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package. *Chaos*, **9**, 413-435.
- HEYDE, C. C., & DAI, W. (1996). On the robustness to small trends of estimation based on the smoothed periodogram. *Journal of Time Series Analysis*, **17**, 141-150.

- HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika*, **68**, 165-176.
- HOSKING, J. R. M. (1984). Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, **20**, 1898-1908.
- HUITEMA, B. E., & MCKEAN, J. W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Methods*, **3**, 104-116.
- HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770-799.
- IVANOV, P. C., NUNES AMARAL, L. A., GOLDBERGER, A. L., & STANLEY, H. E. (1998). Stochastic feedback and the regulation of biological rhythms. *Europhysics Letters*, **43**, 363-368.
- JENKINS, G. M., & WATTS, D. G. (1968). *Spectral analysis and its applications*. San Francisco: Holden-Day.
- JENSEN, H. J. (1998). *Self-organized criticality*. Cambridge: Cambridge University Press.
- JENSEN, H. J., CHRISTENSEN, K., & FOGEDBY, H. C. (1989). $1/f$ noise, distribution of lifetimes, and a pile of sand. *Physical Review B*, **40**, 7425-7427.
- KANTZ, H., & SCHREIBER, T. (1997). *Nonlinear time series analysis*. Cambridge: Cambridge University Press.
- KAPLAN, D. T. (1997). Nonlinearity and nonstationarity: The use of surrogate data in interpreting fluctuations. In M. Di Rienzo, G. Mancia, G. Parati, A. Pedotti, & A. Zanchetti (Eds.), *Frontiers of blood pressure and heart rate analysis*. Amsterdam: IOS.
- KAPLAN, D. T. (1999). *Notes for the 1999 IEEE-EMBS summer school on advanced biomedical signal processing: Nonlinear function estimation, surrogate data, 1/f noise*. Retrieved December 1, 2001 from <http://www.macalester.edu/~kaplan/preprints/welcome.html>.
- KASDIN, N. J. (1995). Discrete simulation of colored noise and stochastic processes and $1/f^\alpha$ power law noise generation. *Proceedings of the IEEE*, **83**, 802-827.
- KAULAKYS, B., & MESKAUSKAS, T. (1998). Modeling $1/f$ noise. *Physical Review E*, **58**, 7013-7019.
- KELLY, A., HEATHCOTE, A., HEATH, R., & LONGSTAFF, M. (2001). Response time dynamics: Evidence for linear and low-dimensional nonlinear structure in human choice sequences. *Quarterly Journal of Experimental Psychology A*, **54**, 805-840.
- KELSO, J. A. S., FUCHS, A., HÖLROYD, T., LANCASTER, R., CHEYNE, D., & WEINBERG, H. (1998). Dynamic cortical activity in the human brain reveals motor equivalence. *Nature*, **392**, 814-818.
- KETTUNEN, J., & KELITKANGAS-JÄRVINEN, L. (2001). Smoothing enhances the detection of common structure from multiple time series. *Behavior Research Methods, Instruments, & Computers*, **33**, 1-9.
- KIM, C.-J., & NELSON, C. R. (1999). *State-space models with regime switching*. Cambridge, MA: MIT Press.
- KÜNSCH, H. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability*, **23**, 1025-1030.
- LAMING, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.
- LAMING, D. R. J. (1979a). Autocorrelation of choice-reaction times. *Acta Psychologica*, **43**, 381-412.
- LAMING, D. R. J. (1979b). Choice reaction performance following an error. *Acta Psychologica*, **43**, 199-224.
- LAMING, D. R. J. (1979c). A critical comparison of two random-walk models for two-choice reaction time. *Acta Psychologica*, **43**, 431-453.
- LO, A. W. (1991). Long term memory in stock market prices. *Econometrica*, **59**, 1279-1313.
- LOBATO, I. N., & SAVIN, N. E. (1998). Real and spurious long memory properties of stock market data. *Journal of Business & Economic Statistics*, **16**, 261-267.
- MADISON, G. (2001). Variability in isochronous tapping: Higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 411-422.
- MANDELBROT, B. B. (1977). *Fractals: Form, chance, and dimension*. San Francisco: Freeman.
- MANDELBROT, B. B. (1982). *The fractal geometry of nature*. San Francisco: Freeman.
- MANDELBROT, B. B., & VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, 422-437.
- MARINARI, E., PARISI, G., RUELLE, D., & WINDEY, P. (1983). Random walk in a random environment and $1/f$ noise. *Physical Review Letters*, **50**, 1223-1225.
- MAYLOR, E. A., CHATER, N., & BROWN, G. D. A. (2001). Scale invariance in the retrieval of retrospective and prospective memories. *Psychonomic Bulletin & Review*, **8**, 162-167.
- MERRILL, W. J., JR., & BENNETT, C. A. (1956). The application of temporal correlation techniques in psychology. *Journal of Applied Psychology*, **40**, 272-280.
- MILOTTI, E. (1995). Linear processes that produce $1/f$ or flicker noise. *Physical Review E*, **51**, 3087-3103.
- MYUNG, I. J. (2002). *Maximum likelihood estimation*. Manuscript submitted for publication.
- MYUNG, I. J., FORSTER, M. R., & BROWNE, M. W. (Eds.) (2000). Model selection [Special issue]. *Journal of Mathematical Psychology*, **44**(1).
- NEWELL, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In William G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.
- NINNESS, B. (1998). Estimation of $1/f$ noise. *IEEE Transactions on Information Theory*, **44**, 32-46.
- NOVIKOV, E., NOVIKOV, A., SHANNAHOFF-KHALSA, D., SCHWARTZ, B., & WRIGHT, J. (1997). Scale-similar activity in the brain. *Physical Review E*, **56**, R2387-R2389.
- OOMS, M., & DOORNIK, J. A. (1999). *Inference and forecasting for fractional autoregressive integrated moving average models with an application to US and UK inflation* (Report EI-9947/A). Rotterdam: Erasmus University, Econometric Institute.
- PAGANO, M. (1974). Estimation of models of autoregressive signal plus white noise. *Annals of Statistics*, **2**, 99-108.
- PHILPOTT, S. J. F. (1934). A theoretical curve of fluctuations of attention. *British Journal of Psychology*, **25**, 221-255.
- PILGRAM, B., & KAPLAN, D. T. (1998). A comparison of estimators for $1/f$ noise. *Physica D*, **114**, 108-122.
- PLOEGER, A., VAN DER MAAS, H. L. J., & HARTELMAN, P. A. I. (2002). Stochastic catastrophe analysis of switches in the perception of apparent motion. *Psychonomic Bulletin & Review*, **9**, 26-42.
- PORT, R. F., & VAN GELDER, T. (Eds.) (1995). *Mind as motion*. Cambridge, MA: MIT Press.
- PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A., & VETTERLING, W. T. (1986). *Numerical recipes*. Cambridge: Cambridge University Press.
- PRESSING, J. (1999a). The referential dynamics of cognition and action. *Psychological Review*, **106**, 714-747.
- PRESSING, J. (1999b). Sources for $1/f$ noise effects in human cognition and performance. In R. Heath, B. Hayes, A. Heathcote, & C. Hooker (Eds.), *Proceedings of the 4th Conference of the Australasian Cognitive Science Society*. Newcastle, NSW: Newcastle University Press.
- PRESSING, J., & JOLLEY-ROGERS, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics*, **76**, 339-347.
- PRIESTLEY, M. B. (1981). *Spectral analysis and time series*. London: Academic Press.
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- RATCLIFF, R., & ROUDER, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, **9**, 347-356.
- RATCLIFF, R., & SMITH, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, **111**, 333-367.
- RATCLIFF, R., VAN ZANDT, T., & MCKOON, G. (1999). Comparing connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.
- ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, **23**, 1048-1072.
- ROE, R. M., BUSEMEYER, J. R., & TOWNSEND, J. T. (2001). Multi-alternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, **108**, 370-392.
- SAMORODNITSKY, G., & TAQQU, M. S. (1993). Linear models with long-range dependence and with finite or infinite variance. In D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, & M. S. Taquq (Eds.), *New directions in time series analysis: Part II* (pp. 325-373). New York: Springer-Verlag.

- SAUPE, D. (1988). Algorithms for random fractals. In H. Peitgen & D. Saupe (Eds.), *The science of fractal images* (pp. 71-136). New York: Springer-Verlag.
- SCHEPERS, H. E., VAN BEEK, J. H. G. M., & BASSINGTHWAIGHTE, J. B. (1992). Four methods to estimate the fractal dimension from self-affine signals. *IEEE Engineering in Medicine & Biology*, **11**, 57-71.
- SCHREIBER, T., & SCHMITZ, A. (2000). Surrogate time series. *Physica D*, **142**, 346-382.
- SCHROEDER, M. (1991). *Fractals, chaos, power laws: Minutes from an infinite paradise*. New York: Freeman.
- SHEU, C.-F., & RATCLIFF, R. (1995). The application of Fourier deconvolution to reaction time data: A cautionary note. *Psychological Bulletin*, **118**, 285-299.
- SMITH, A. (2002). *Why regime switching creates the illusion of long memory*. Retrieved March 1, 2002 from http://www.agecon.ucdavis.edu/faculty_pages/smith/working-papers/LMandBreaks.pdf.
- SMITH, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time. *Psychological Bulletin*, **108**, 533-550.
- SNODGRASS, J. G., LUCE, R. D., & GALANTER, E. (1967). Some experiments on simple and choice reaction time. *Journal of Experimental Psychology*, **75**, 1-17.
- SOWELL, F. B. (1992a). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53**, 165-188.
- SOWELL, F. B. (1992b). Modeling long run behavior with the fractional ARIMA model. *Journal of Monetary Economics*, **29**, 277-302.
- STUART, A., & ORD, J. K. (1991). *Kendall's advanced theory of statistics* (Vol. 2). New York: Oxford University Press.
- TAQQU, M. S., & TEVEROVSKY, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance time series. In R. Adler, R. Feldman, & M. S. Taqqu (Eds.), *A practical guide to heavy tails: Statistical techniques and applications* (pp. 177-217). Boston: Birkhauser.
- TAQQU, M. S., TEVEROVSKY, V., & WILLINGER, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals*, **3**, 785-798.
- TEVEROVSKY, V., TAQQU, M. S., & WILLINGER, W. (1999). A critical look at Lo's modified R/S statistic. *Journal of Statistical Planning & Inference*, **80**, 211-227.
- THELEN, E., & SMITH, L. B. (1994). *A dynamical systems approach to the development of cognition and action*. Cambridge, MA: Bradford Books/MIT Press.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- TOWNSEND, J. T., HU, G. G., & KADLEC, H. (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. *Perception & Psychophysics*, **43**, 575-591.
- USHER, M., & MCCLELLAND, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, **108**, 550-592.
- USHER, M., STEMMLER, M., & OLAMI, Z. (1995). Dynamic pattern formation leads to $1/f$ noise in neural populations. *Physical Review Letters*, **74**, 326-329.
- VAN DER MAAS, H., DOLAN, C., & MOLENAAR, P. (2002). *Phase transitions in the tradeoff between speed and accuracy in choice reaction time tasks*. Manuscript submitted for publication.
- VAN DER ZIEL, A. (1950). On the noise spectra of semiconductor noise and flicker effects. *Physica*, **16**, 359-372.
- VAN ORDEN, G. C., MORENO, M. A., & HOLDEN, J. G. (2003). A proper metaphysics for cognitive performance. *Nonlinear Dynamics, Psychology, & Life Sciences*, **7**, 49-60.
- VAN ORDEN, G. C., PENNINGTON, B. F., & STONE, G. O. (2001). What do double dissociations prove? *Cognitive Science*, **25**, 111-172.
- VAN ZANDT, T., COLONIUS, H., & PROCTOR, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, **7**, 208-256.
- VERPLANCK, W. S., COLLIER, G. H., & COTTON, J. W. (1952). Nonindependence of successive responses in measurements of the visual threshold. *Journal of Experimental Psychology*, **44**, 273-281.
- VOSS, R. F. (1988). Fractals in nature: From characterization to simulation. In H. Peitgen & D. Saupe (Eds.), *The science of fractal images* (pp. 21-70). New York: Springer-Verlag.
- WAGENMAKERS, E.-J., & FARRELL, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, **11**, 192-196.
- WARD, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.
- WEISS, B., COLEMAN, P. D., & GREEN, R. F. (1955). A stochastic model for the time-ordered dependencies in continuous scale repetitive judgements. *Journal of Experimental Psychology*, **50**, 237-244.
- WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, **9**, 60-62.
- WIXTED, J. T., & EBBESEN, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, **25**, 731-739.
- WOLD, H. O. A. (1938). *A study in the analysis of stationary time series*. Uppsala: Almqvist & Wiksell.
- WOODROW, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, **13**, 473-499.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology*. New York: Holt.

NOTES

1. The Internet source <http://www.nslj-genetics.org/wli/1fnoise/> gives an ordered summary of the literature and provides many more examples that attest to the ubiquity of $1/f$ noise.

2. Analyses are often carried out on residual RTs that can be obtained by subtracting the mean RT for each condition from the corresponding observations—that is, on the error variance that remains after the treatment effects have been removed from the data.

3. Another, less popular definition of a long-range process is that the variance of its sample mean decreases at a rate slower than n^{-1} (Beran, 1994).

4. It is of interest that a description in the frequency domain and one in the ACF are related through the Wiener-Khinchin theorem. From this celebrated theorem, it follows that the normalized power spectral density function, $S(f)$, is the Fourier transform of the autocorrelation function, $C(k)$:

$$S(f) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T C(k) \cos(2\pi fk) dk.$$

This means that Figures 2 and 3 (ACFs, time domain) contains the same information as Figures 4 and 5 (power spectra, frequency domain), and one can be obtained from the other through transformation.

5. Equation 14 implies that $C(k)$ and $m_c k^{-\alpha}$ are asymptotically equal (see, e.g., Apostol, 1966, p. 396), a relation that can be symbolically indicated by writing $C(k) \sim m_c k^{-\alpha}$ as $n \rightarrow \infty$.

6. This was pointed out to us by Jerry Busemeyer.

7. This result finds practical application in the regulations of the International Association of Athletics Federations (IAAF), which state that a runner is to be disqualified if he or she leaves the starting position before 100 msec have elapsed after the starting signal.

8. The ARFIMA package, written in the matrix programming language Ox (Doornik, 2001), is available at <http://www.nuff.ox.ac.uk/Users/Doornik/>, and the R fracdiff package, maintained by Friedrich Leisch, can be obtained from <http://cran.r-project.org>.

9. Quadratic detrending of the series flattens the lowest frequencies, and this is inconsistent with the ARFIMA(1, d , 1) model. Hence, such preprocessing made it more difficult to distinguish between the competing models. The issues of nonstationarity, $1/f$ noise, and detrending is in need of further study.

10. Although not shown, the accuracy rates and RT distributions for the sequential sampling model reflected those found in the choice RT experiment.

APPENDIX

Fractional Differencing and the ARFIMA Model

The ARFIMA(p, d, q) model consists of an ARMA(p, q) component for the quantification of short-range processes and a fractional differencing parameter d for the quantification of a long-range process. Here, we first describe the process of fractional differencing in the absence of any short-range processes [i.e., ARFIMA(0, d , 0) or fractionally differenced white noise] and then briefly consider the more complete model.

Recall that differencing a random walk series, say, X_t , results in a white noise series, ε_t . This can be formulated as

$$\nabla X_t = (1 - B)X_t = \varepsilon_t, \quad (\text{A1})$$

where ∇ is the differencing operator and B is the backward shift operator defined as $BX_t = X_{t-1}$ (see Equation 3). Thus,

$$X_t - X_{t-1} = (1 - B)X_t, (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = (1 - B)^2 X_t,$$

and so on. In the case in which d can take on integer values only, incorporating ∇ in the ARMA(p, q) model yields the traditional Box and Jenkins (1970) ARIMA(p, d, q) time series models. For instance, ARIMA(0, 0, 0) is a white noise process and ARIMA(0, 1, 0) is a random walk process. The ARIMA approach can be extended by allowing a series to be differenced d times, where d can take on real values. Using binomial series expansion (see, e.g., Apostol, 1966, p. 441), the fractional differencing operator ∇^d is given by

$$\nabla^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k,$$

with binomial coefficients

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)},$$

where $\Gamma(\cdot)$ is the gamma function (Beran, 1994). Thus, applying fractional differencing to a time series X_t is described as

$$\begin{aligned} \nabla^d X_t &= (1 - B)^d X_t = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k X_t = \\ X_t - dX_{t-1} + \frac{1}{2}d(d-1)X_{t-2} - \frac{1}{6}d(d-1)(2-d)X_{t-3} + \dots \end{aligned} \quad (\text{A2})$$

(Hosking, 1984). If $\nabla^d X_t$ is a white noise process (i.e., $\nabla^d X_t = \varepsilon_t$ or $X_t = \nabla^{-d} \varepsilon_t$, where ε_t is a purely random process), then X_t is an ARFIMA(0, d , 0) process and is called *fractionally differenced white noise*. X_t is stationary provided that $-1/2 < d < 1/2$. If $d > 0$, X_t is long-range dependent with an ACF whose asymptotic behavior is given by

$$C(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \quad k \rightarrow \infty \quad (\text{A3})$$

and for the spectral density,

$$S(\omega) \sim \omega^{-2d} \quad \omega \rightarrow 0 \quad (\text{A4})$$

For processes with finite variance, $d \in (0, 1/2)$ is related to self-similarity parameter H (Hurst, 1951; Mandelbrot & Van Ness, 1968) as $d = H - 1/2$ (Taqq & Teverovsky, 1998). Also, if the spectrum is given by $1/f^\alpha$, $d = 1/2\alpha$ —that is, in the log-log power spectrum d is half the (positive) value of the spectral slope. The ARFIMA(0, d , 0) process that is long-range dependent and stationary for $d \in (0, 1/2)$ can be extended to incorporate short-range processes [i.e., ARMA(p, q) processes]. The complete ARFIMA(p, d, q) process may be described as a time series X_t , satisfying

$$\begin{aligned} w_t &= \nabla^d (X_t - \mu) \\ w_t - \phi_1 w_{t-1} - \dots - \phi_p w_{t-p} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \end{aligned} \quad (\text{A5})$$

where μ is the series mean (Hosking, 1984). An alternative and more concise formula is to define the ARFIMA(p, d, q) model as

$$\Phi(B)(1 - B)^d (X_t - \mu) = \Theta(B)\varepsilon_t, \quad (\text{A6})$$

where $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ specify the AR and MA polynomials, respectively (see, e.g., Beran, 1994), and ε_t is a purely random process with $E(\varepsilon_t) = 0$ and variance σ_ε^2 . For sta-

APPENDIX (Continued)

tionarity of the ARMA part of the model, it is further required that the roots of $\Phi(z) = 0$ and $\Theta(z) = 0$ lie outside the unit circle (see, e.g., Priestley, 1981, pp. 132–135). As was noted by Beran (1994), the ARFIMA(p, d, q) process might be conceptualized as fractionally differenced white noise [i.e., ARFIMA(0, $d, 0$)] passed through an ARMA(p, q) filter. The ACF and spectral density of the ARFIMA(p, d, q) process are asymptotically equal to the ACF and spectral density of the ARFIMA(0, $d, 0$) process, since the behavior at high lags (low frequencies) is determined exclusively by the long-range parameter d . Further details on the ARFIMA(p, d, q) process and the exact Gaussian maximum likelihood procedure used here to estimate its parameters can be found elsewhere (e.g., Baillie, 1996; Beran, 1994; Doornik & Ooms, 2003; Hosking, 1981, 1984; Ooms & Doornik, 1999; Sowell, 1992a, 1992b).

(Manuscript received July 12, 2002;
revision accepted for publication July 1, 2003.)