Journal of Experimental Psychology: General
2005, Vol. 134, No. 1, 108–116

0096-3445/05/$12.00 DOI: 10.1037/0096-3445.134.1.108

COMMENTS

Human Cognition and a Pile of Sand: A Discussion on Serial Correlations and Self-Organized Criticality

Eric-Jan Wagenmakers
University of Amsterdam

Simon Farrell
University of Bristol

Roger Ratcliff
The Ohio State University

Recently, G. C. Van Orden, J. G. Holden, and M. T. Turvey (2003) proposed to abandon the conventional framework of cognitive psychology in favor of the framework of nonlinear dynamical systems theory. Van Orden et al. presented evidence that "purposive behavior originates in self-organized criticality" (p. 333). Here, the authors show that Van Orden et al.'s analyses do not test their hypotheses. Further, the authors argue that a confirmation of Van Orden et al.'s hypotheses would not have constituted firm evidence in support of their framework. Finally, the absence of a specific model for how self-organized criticality produces the observed behavior makes it very difficult to derive testable predictions. The authors conclude that the proposed paradigm shift is presently unwarranted.

In a provocative article, Van Orden, Holden, and Turvey (2003) recommended the framework of nonlinear dynamical systems theory as an attractive alternative to the current mainstream paradigms in cognitive psychology. Van Orden et al. used time series analysis to confirm their hypothesis that fluctuations in performance over the course of an experiment display persistent serial correlations or so-called $1/f^{\alpha}$ noise. The presence of $1/f^{\alpha}$ noise was then taken as evidence for the important role of "self-organized criticality" (Van Orden et al., p. 333) in human cognition. Van Orden et al. went on to conclude that "intentional acts originate in states of self-organized criticality" (p. 347) and further claimed that "old science" (p. 347; i.e., the set of current paradigms in cognitive psychology) is fundamentally flawed.

We believe that the evidence for questioning the conventional paradigms of cognitive psychology is not as strong as claimed by Van Orden et al. (2003). After all, psychologists practicing traditional cognitive psychology have not usually attempted to explain serial correlations. In fact, the dominant approach to the study of human cognition typically ignores serial dependence, deems it irrelevant, or treats it as a nuisance variable (e.g., Gilden, 2001; Slifkin & Newell, 1998; cf. also Van Orden et al., 2003). When new scientific evidence overthrows an existing dominant research paradigm, a strict requirement should be that this new evidence is sound (i.e., sufficiently replicated) and that the old research paradigm fails to account for the new findings even after considerable effort is expended to reconcile the problematic findings with present theory. In addition, a new paradigm should only be accepted when it is shown to naturally account for existing findings that were problematic for the old paradigm (in addition to providing an adequate account for many of the findings that the old paradigm handled adequately). We show in this comment that none of these three conditions have been met for the evidence presented by Van Orden et al.

Although we find the application of self-organized criticality (SOC) to human cognition conceptually intriguing and even innovative and promising, it is important to objectively and critically assess the contribution of the new paradigm. We offer two main criticisms with respect to the research and theoretical claims advanced by Van Orden et al. (2003). Future research will have to show whether these criticisms can be satisfactorily addressed. First, Van Orden et al. avoided any effort to reconcile their findings with "old science." Instead, they dismiss traditional "component-dominant" (p. 335) accounts for their data before the fact. We believe this is a serious omission, and we demonstrate that Van Orden et al. made this omission both at the level of statistics and at the level of theory. If the new science of nonlinear dynamical systems theory is truly superior to the current paradigm of cognitive psychology, then its proponents should not shy away from a head-on quantitative comparison. If, however, the new paradigm turns out not to be superior, then there is little motivation for abandoning the current framework, which has, after all, a proven track record.

Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands; Simon Farrell, Department of Experimental Psychology, University of Bristol, Bristol, United Kingdom; Roger Ratcliff, Department of Psychology, The Ohio State University.

Preparation of this article was supported by National Institute of Mental Health Grants R37-MH44640 and K05-MH01891. We thank Amy Criss, Marie Jones, Philip Smith, and René Zeelenberg for comments on a previous version of this article, and we are grateful to Mingzhou Ding, Guy C. Van Orden, and Lawrence Ward for their suggestions for improvement. We also thank Guy C. Van Orden, John G. Holden, and Michael T. Turvey for making their data available to us.

Correspondence concerning this article should be addressed to Eric-Jan Wagenmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, the Netherlands. E-mail: ewagenmakers@fmg.uva.nl

Second, Van Orden et al. (2003) supported their hypothesis in favor of nonlinear dynamical systems mostly with metaphors such as heating oil in a pan or juggling soap. Although such metaphors are persuasive in the context of the more general issues of intentional control and provide useful everyday analogues to the complex conceptual issues, it is not clear to us how a verbal description about juggling soap can lead to the construction of an appropriate quantitative model for human cognition. In an influential book on SOC, Jensen (1998) stated that verbal descriptions and metaphors of SOC in the absence of concrete models are "rather abstract, heuristic wishful thinking" (p. 6). The principle of SOC should be evaluated on a case-by-case basis, using clearly specified quantitative models (cf. Gisiger, 2001). The need for specific models is underscored by the fact that SOC is acutely sensitive to the details of the dynamics of the system under study (e.g., Gisiger, 2001; Jensen, 1998), as is illustrated later.

The outline of this comment is as follows. First, we clarify the fundamental issues raised by Van Orden et al. (2003), and we demonstrate that the reported analyses are not very informative and do not test what the authors had set out to examine. We outline how a diagnostic test can be conducted (cf. Beran, 1994; for an application to psychology, see Wagenmakers, Farrell, & Ratcliff, 2004), and we then apply this test to the data of the two experiments presented by Van Orden et al. The results of this test provide weak support for Van Orden et al.'s assertion of $1/f^{\alpha}$ noise. Next, we show that the finding of $1/f^{\alpha}$ noise does not constitute firm evidence for SOC. We also briefly discuss and defend a simple alternative explanation for $1/f^{\alpha}$ noise from the field of econometrics (Granger, 1980), although other plausible alternative explanations certainly exist. In addition, we argue that, contrary to the claim by Van Orden et al., sequential sampling models can be linked to specific biological processes, as illustrated by recent findings in neuroscience. Finally, we argue that the proposed concept of SOC can only be scientifically useful when it is embedded in a specific model. Specific models make testable predictions, and this is a hallmark of good science (Roberts & Pashler, 2000). We conclude that the proposed new framework for human cognition, although novel and interesting, is, at this point in its development, quite seriously underspecified.

SOC in Human Cognition

Van Orden et al. (2003) advanced the opinion that human cognition and behavior show SOC (e.g., Bak, 1996; Sornette, 2000; Ward, 2002). This concept is perhaps best explained by considering the classic example of SOC: a pile of sand (see Jensen, 1998, for further details). Specifically, consider a system that consists of grains of sand piled up in a corner (i.e., bounded by two orthogonal edges or walls). At random positions along the edges, new grains of sand are dropped onto the pile one by one. When the local slope of the sand pile exceeds a certain threshold (i.e., it is sufficiently steep), grains of sand will be transported downhill until the local slope is again below threshold. As a result of this mechanism, avalanches of different sizes can occur; if several adjacent slopes happen to be near threshold, a single grain of sand may be sufficient to cause a cascade of avalanches. If a grain of sand is transported all the way down to the foot of the pile, it is removed from the system (e.g., imagine the sand pile system positioned on a table with grains of sand at the bottom of the pile falling off the edges of the table).

The above pile of sand can be said to self-organize to reach a critical state. Once the pile is in this critical state, small perturbations (i.e., single grains of sand added to the pile) can sometimes have dramatic consequences (i.e., large avalanches). Models based on similar principles have also been applied to evolution (e.g., Bak & Sneppen, 1993; but see Davidsen & Lüthje, 2001), forest fires (e.g., Malamud, Morein, & Turcotte, 1998), and earthquakes (e.g., Davidsen & Paczuski, 2002; Davidsen & Schuster, 2000, 2002) and go under the generic label of SOC (for an overview, see Paczuski, Maslov, & Bak, 1996). For SOC to be present, a system needs to be gradually pushed toward a threshold, and, in addition, there need to be dominant interactions between many degrees of freedom or individual units. Hence, Jensen (1998) termed these kinds of models "*slowly driven, interaction-dominated threshold* systems" (p. 126).

Van Orden et al. (2003) claimed that human cognition behaves just as piles of sand, evolution, and earthquakes do. That is, Van Orden et al. hypothesized that the dynamics of these very diverse systems share fundamental principles with human cognition and that considering human cognition in terms of these systems is more theoretically meaningful than standard "component dominant" (see Van Orden et al., 2003, p. 335) approaches to cognition. We believe that this theory deserves serious attention, if only because of its generality and scientific appeal. For example, Bak and colleagues (e.g., Bak & Chialvo, 2001; Chialvo & Bak, 1999) have recently highlighted the adaptive nature of self-organizing neural networks: A network that is in a state of criticality is able to quickly reorganize and swiftly adapt to new situations (Alstrøm & Stassinopoulos, 1995; cf. Linkenkaer-Hansen, 2002; Linkenkaer-Hansen, Nikouline, Palva, & Ilmoniemi, 2001).

Traditional methods in psychology are not well suited to test whether the workings of human cognition relate to the behavior of a pile of sand. However, Van Orden et al. (2003) noted that *time series analysis* of cognitive performance, or the analysis of how cognitive performance fluctuates over the course of an experiment, might be able to reveal aspects of human behavior supporting the pile of sand hypothesis of human cognition. In particular, under certain conditions, for certain dependent variables, SOC systems display persistent serial correlations or $1/f^{\alpha}$ noise. In fact, the notion of SOC was first introduced to explain why $1/f^{\alpha}$ noise occurs in so many quite different natural systems (Bak, Tang, & Wiesenfeld, 1987; but see Jensen, 1998, and Jensen, Christensen, & Fogedby, 1989; for a recent review with respect to biological systems, see Gisiger, 2001).

Motivated by this pile of sand hypothesis, Van Orden et al. (2003) set out to determine whether the temporal fluctuations in human cognition show $1/f^{\alpha}$ noise (cf. Gilden, 1997, 2001; Gilden, Thornton, & Mallon, 1995), consistent with the predictions from SOC. The pattern of serial correlations known as $1/f^{\alpha}$ noise is special not just because it occurs often in all kinds of natural systems but also because only particular types of models appear to produce $1/f$ noise.[1]

$1/f^{\alpha}$ Noise and How to Detect It

Most psychological research on $1/f^{\alpha}$ noise proceeds globally as follows. First, the temporal order in which the observations have

[1] See http://www.nslij-genetics.org/wli/1fnoise/ for an ordered summary of the scientific literature on $1/f^{\alpha}$ noise.

been collected is kept intact. Often, condition means are subtracted from each observation to reduce the impact of task difficulty on the trial-by-trial fluctuations in performance. The resulting series is then analyzed in the frequency domain (for an introduction of time series analysis, see Priestley, 1981); this means that the temporal fluctuations are reexpressed as an infinite sum of sine and cosine terms, each with a specified frequency and amplitude. For example, for a time series that can be characterized by pronounced slow waves, its power (i.e., squared amplitude) will concentrate mostly at the low frequencies. The results of the analysis in the frequency domain are typically displayed by plotting frequency against squared amplitude on log-log axes. In this so-called log-log power spectrum, a $1/f^{\alpha}$ noise process ideally produces a straight line with slope $-\alpha$. Thus, a best fitting straight line is drawn through the spectrum, and the slope of this line estimates the intensity of the $1/f^{\alpha}$ noise process, where α usually ranges from 0.5 to 1.5. This is the first method that Van Orden et al. (2003) used (for an extended discussion of this method, see Wagenmakers et al., 2004).

The immediate problem with this procedure is that there is no statistical test to determine whether $1/f^{\alpha}$ noise is present. Although fitting a regression line to a log-log power spectrum may yield a negative slope, this does not imply that the data follow a straight line. The reader is left to judge, by eye, whether he or she believes the straight line produces an accurate fit. A second, related problem is that no alternative models are considered. This is particularly worrisome when one considers the qualitative manner in which $1/f^{\alpha}$ noise differs from the kind of temporal fluctuations generated by very simple models. For a $1/f^{\alpha}$ noise process, serial correlations (i.e., the correlations between ordered trials) decay very slowly with the number of intervening trials (i.e., persistent serial correlations). Specifically, the correlation C with k intervening trials is given by $C(k) = |k|^{-\gamma}$, with γ between 0 and 1. Thus, the decay of serial correlations follows a power function. In contrast, transient serial correlations decay fairly quickly with the number of intervening trials, for instance, as an exponential function. Such transient correlations are easily generated by standard time series models such as the class of autoregressive moving average (ARMA) models, in which behavior at time t is related to behavior at time $t - k$ via a procedure similar to linear regression: $X_t = \sum_{r=1}^{p} \phi_r X_{t-r} + \varepsilon_t + \sum_{r=1}^{q} \theta_r \varepsilon_{t-r}$, where ε is white noise and p and q determine the order of the ARMA process. The foregoing implies that the difference between $1/f^{\alpha}$ noise and standard, readily explained processes that yield transient correlations can also be formulated as the difference between power function decay and, for example, exponential function decay (Beran, 1994). Thus, the difference between a persistent $1/f^{\alpha}$ noise process and a transient ARMA process is not in the absolute value of serial correlations but in the rate of decay of these serial correlations with increasing lag. These different patterns of serial correlation map directly onto the power spectrum (i.e., via the Wiener–Khinchin theorem; e.g., Priestley, 1981) such that a $1/f^{\alpha}$ noise process follows a straight line in the power spectrum, whereas a transient process flattens at the lower frequencies. The leveling off at the lower frequencies indicates that there are no correlations between trials that are spaced widely apart.

We are now in a position to discuss the analyses and the data from Van Orden et al. (2003) in more detail. In Van Orden et al. (2003), the top right panel of their Figure 1 (p. 336) plots the log-log power spectrum together with the best fitting straight line for an example subject from their Experiment 1 (i.e., simple response time; RT). The slope of this line is about -0.6. Can one infer from this, as Van Orden et al. did, that the underlying process is indeed a $1/f^{\alpha}$ noise process with α about 0.6? The answer is emphatically no. Rather than testing the hypothesis of $1/f^{\alpha}$ noise, the above procedure already assumes that the underlying process is a $1/f^{\alpha}$ noise process. If the authors assumed a different kind of underlying process was at work in this situation, then it would make no sense to fit a straight line to the spectrum, because the log-log spectra of many different kinds of processes are not linear. In fact, close examination of Van Orden et al.'s Figure 1 suggests that their observed spectrum may level off at the low frequencies, a phenomenon indicative of a simple ARMA process, not a $1/f^{\alpha}$ noise process. Although our perception might be mistaken, this highlights the case that, without quantitative assessment of alternatives, the nature of power spectra is open to alternative interpretation.

To be fair, we must mention that the authors did perform one statistical test: They tested whether the serial correlations (quantified, rightly or wrongly, by the spectral slopes) differed significantly from a white noise process without any serial correlations. It comes as no surprise that the null hypothesis of no serial correlation was confidently rejected. Of course, the null hypothesis would also have been rejected if the data came from a standard ARMA process that produces transient serial correlations. Hence, this analysis does not address the issue of whether the observed data are generated by a $1/f^{\alpha}$ noise process or an ARMA process (cf. Rangarajan & Ding, 2000, p. 4995). The very same comments apply to the spectral analysis of Experiment 2 (word reading), except that the reported slopes are so shallow that they fall outside of the range of what is typically considered a $1/f$ noise process (Beran, 1994). An important reason for disregarding shallow slopes is that because the underlying processes are stochastic in nature, such a pattern is extremely hard to distinguish from an ARMA pattern of transient correlations.

In sum, the spectral analyses by Van Orden et al. (2003) convincingly demonstrated the presence of serial correlations.[2] Although it may be of scientific interest (Laming, 1968), this general finding does not confirm Van Orden et al.'s hypothesis. Van Orden et al. wanted to test whether the data are consistent with a $1/f^{\alpha}$ noise process. Their analyses do not address this issue, because a process that yields only transient correlations (e.g., an ARMA process) may have been responsible for generating their data.

In a second analysis reported by Van Orden et al. (2003), they plotted variance as a function of sample size. Like the spectral slope measure, such a dispersion analysis is a heuristic measure (Beran, 1994). For the kind of sample sizes used in psychology, transient ARMA processes may spuriously affect the best fitting slope of a dispersion analysis (e.g., Rangarajan & Ding, 2000). Again, for such an analysis, a test against the white noise null

[2] Note that the serial correlations in Van Orden et al.'s (2003) simple RT experiment are much more pronounced than those reported in Gilden et al. (1995) and Wagenmakers et al. (2004). This might be due to the fact that Van Orden et al. used fixed intervals that made the onset of the stimulus completely predictable. This may effectively turn the experiment into a temporal estimation task, which is known to result in relatively high serial correlations (Gilden, 2001; Wagenmakers et al., 2004).

hypothesis does show that serial correlations are present, but it does not reveal the nature of these serial correlations.

The above discussion highlights the need to test the hypothesis of a $1/f^{\alpha}$ noise process versus the hypothesis of an alternative process that incorporates serial correlations other than that of the $1/f^{\alpha}$ type. An ideal candidate for such an alternative process is the standard ARMA process that generates transient correlations. A test between the $1/f^{\alpha}$ noise model and the ARMA model may be accomplished by autoregressive fractionally integrated moving average (ARFIMA) time series modeling, popular in statistics and economics (e.g., Baillie, 1996; Beran, 1994; Wagenmakers et al., 2004). The fractional integration in ARFIMA models is incorporated as a single parameter *d* that is associated with a $1/f^{\alpha}$ noise process. ARFIMA models are a generalization of ARMA models and are able to describe both transient and persistent serial correlations. Model selection techniques (e.g., Burnham & Anderson, 2002; Myung, Forster, & Browne, 2000; Wagenmakers & Farrell, 2004) can then be used to determine which members of the ARFIMA family are the most plausible (cf. Farrell, Wagenmakers, & Ratcliff, 2004). A detailed discussion about ARFIMA models can be found elsewhere (e.g., Baillie, 1996; Beran, 1994; Hosking, 1981, 1984). In the next section, we apply the ARFIMA modeling technique to the data presented in Van Orden et al. (2003).

ARFIMA Modeling of the Van Orden et al. (2003) Data

In their article, Van Orden et al. (2003) presented the results of two experiments: the first, a simple RT experiment with 10 participants; the second, a word-reading (or word-naming) experiment with 20 participants. The ARFIMA analyses of these data are based on the preprocessed individual time series. Preprocessing involved detrending and standardizing (cf. Van Orden et al., 2003). Each individual time series consisted of 1,024 observations.

To each individual time series from Van Orden et al. (2003), we fitted 18 models using maximum likelihood estimation as implemented in the Ox ARFIMA software package (Doornik, 2001; Doornik & Ooms, 2003; Ooms & Doornik, 1999). Nine of the models are ARMA models and hence do not contain long-range serial correlations or $1/f^{\alpha}$ noise. These ARMA models differ only in the number of autoregressive (AR) and moving average (MA) parameters. The number of parameters was systematically varied (see Table 1), ranging from a model without any serial dependence—that is, ARMA (0, 0)—to an ARMA (2, 2) model. To illustrate, the ARMA (2, 2) model has two AR parameters and two MA parameters: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$. The other 9 models are ARFIMA models and differ from the set of ARMA models by their inclusion of a parameter *d* that represents the persistence of serial correlations.[3]

If the Van Orden et al. (2003) data contain persistent correlations or $1/f^{\alpha}$ noise, the ARFIMA models are expected to outperform the transient ARMA models. Thus, the issue of deciding between transient and persistent processes is reduced to the issue of selecting between ARMA and ARFIMA models (cf. Bisaglia, 2002; Bisaglia & Guégan, 1998; Hosking, 1984, Table 4). In the process of deciding between several competing models (i.e., model selection), it is not appropriate to focus solely on goodness of fit (i.e., the maximum likelihood values). Models with many parameters may spuriously obtain a relatively good fit to data by capturing idiosyncratic noise (overfitting; Myung et al., 2000). Such models will not generalize well to other comparable data sets.

Table 1
Average Model Weights, Average Rank Order, and Number of Rank Order Ones (i.e., Wins) for ARMA and ARFIMA Models Fitted to Van Orden, Holden, and Turvey's (2003) Experiment 1 (i.e., Simple Response Time), Separately for AIC and BIC

	AIC				BIC			
	Weight				Weight			
Model	*M*	*SD*	Rank	Wins	*M*	*SD*	Rank	Wins
Transient correlation								
ARMA (0, 0)	.00	.00	18.0	0	.00	.00	18.0	0
ARMA (1, 0)	.01	.01	13.0	0	**.22**	**.25**	5.0	**3**
ARMA (2, 0)	.06	.07	8.5	**1**	.14	.18	5.4	2
ARMA (0, 1)	.00	.00	16.9	0	.00	.00	16.6	0
ARMA (0, 2)	.00	.00	15.2	0	.00	.01	13.8	0
ARMA (1, 1)	.07	.08	7.5	**1**	.14	.14	**3.4**	1
ARMA (2, 1)	**.09**	**.05**	**5.1**	0	.03	.03	6.7	0
ARMA (1, 2)	.07	.05	6.9	0	.02	.02	7.6	0
ARMA (2, 2)	.08	.08	7.4	**1**	.00	.00	11.9	0
Persistent correlation								
ARFIMA (0, *d*, 0)	.01	.04	13.9	0	**.16**	**.30**	10.4	**2**
ARFIMA (1, *d*, 0)	.06	.07	7.5	2	.13	.16	**3.9**	1
ARFIMA (2, *d*, 0)	.06	.03	6.6	0	.01	.01	7.4	0
ARFIMA (0, *d*, 1)	.04	.06	10.5	1	.11	.19	7.1	1
ARFIMA (0, *d*, 2)	.06	.05	7.8	0	.02	.02	8.7	0
ARFIMA (1, *d*, 1)	.06	.03	7.2	0	.01	.01	7.5	0
ARFIMA (2, *d*, 1)	**.16**	**.15**	**5.2**	**3**	.00	.01	11.1	0
ARFIMA (1, *d*, 2)	.06	.06	7.9	1	.00	.00	12.0	0
ARFIMA (2, *d*, 2)	.09	.06	5.9	0	.00	.00	14.5	0

Note. *Weight* denotes the average model weights (i.e., averaged over participants). *Rank* denotes the average rank order of the models (i.e., averaged over participants). The ARMA and ARFIMA models that are best in terms of average weight, average rank order, and total number of wins are in bold. AIC = Akaike's information criterion; BIC = Bayesian information criterion; ARMA = autoregressive moving average; ARFIMA = autoregressive fractionally integrated moving average.

Hence, model selection requires that goodness of fit be discounted as a function of model complexity.

One popular method for model selection is Akaike's information criterion (AIC; e.g., Akaike, 1974; Burnham & Anderson, 2002), given by $\text{AIC} = -2\ln L + 2k$, where L is the maximum likelihood and k is the number of free parameters. The model with the lowest AIC value is preferred. Thus, AIC discounts goodness of fit as a function of the number of free parameters. Another often-used model-selection method is the Bayesian information criterion (BIC; e.g., Raftery, 1995; Schwarz, 1978), given by $\text{BIC} = -2\ln L + k \ln(n)$, where n is the number of observations (for details, see Kass & Raftery, 1995). Thus, the BIC penalty term incorporates both sample size and the number of free parameters. For the data sets considered here, $n = 1{,}024$, and the BIC penalty term for having one additional parameter, that is, $\ln(1024) \approx 6.93$, is considerably higher than the corresponding AIC penalty term of

[3] In previous work, we specifically compared performance of the transient ARMA (1, 1) model to that of the persistent ARFIMA (1, *d*, 1) model (Wagenmakers et al., 2004). The current approach is more general as it includes a wider range of models.

2. This means that compared with AIC, the BIC will prefer models with fewer free parameters. Because AIC and BIC may give different results, we decided to report both.

The raw information criterion (IC, denoting either AIC or BIC) values can be quite difficult to interpret, and hence we performed a simple transformation that yields model weights, as follows:

$$w_i(\mathrm{IC}) = \exp\{-\tfrac{1}{2}\Delta_i(\mathrm{IC})\}/\sum_{k=1}^{K}\exp\{-\tfrac{1}{2}\Delta_k(\mathrm{IC})\}.$$

In this equation, the subscript i denotes the model under consideration, and $\Delta_i(\mathrm{IC})$ is the difference in the information criterion (IC) between model i and the best IC model, that is, $\Delta_i(\mathrm{IC}) = \mathrm{IC}_i - \min \mathrm{IC}$. The sum in the denominator is over all K candidate models (cf. Wagenmakers & Farrell, 2004). The IC weights provide a measure that allows assessment of the strength of evidence for a model conditional on the set of candidate models and the particular IC used.

For each of the 18 candidate models (i.e., 9 ARMA models and 9 ARFIMA models), we calculated AIC weights and BIC weights separately for each participant in the Van Orden et al. (2003) experiments. Table 1 summarizes the results for Van Orden et al.'s simple RT experiment. For every model, Table 1 shows the mean AIC weights and the mean BIC weights, obtained by averaging model weights over the 10 participants, as well as the standard deviation of these weights. Table 1 also shows the mean rank order of each model (i.e., 1.0 being the best possible value and 18.0 being the worst possible value) and the number of participants for which each model yielded the best AIC or BIC fit.

The results demonstrate that there is a considerable difference between AIC and BIC model selection. AIC prefers relatively complex models, whereas the heavy $\ln(1024) \approx 6.93$ penalty term leads BIC to prefer relatively simple models. Second, according to the AIC, the ARFIMA (2, d, 1) model is best with a model weight $w_{2,d,1}(\mathrm{AIC}) = .16$. This result provides only limited support for the $1/f^{\alpha}$ noise hypothesis, however. The value of .16 is not very high (like probabilities, weights add to 1), and as many as five transient ARMA models have AIC weights that are over one third of the weight for the ARFIMA (2, d, 1) model. In addition, the BIC shows an overall preference for the ARMA (1, 0) model, $w_{1,0}(\mathrm{BIC}) = .22$, although the difference with the best ARFIMA model is arguably small, $w_{0,d,0}(\mathrm{BIC}) = .16$.

It is generally the case that models with high weights have low (i.e., primary) rank orders and relatively high numbers of wins. An exception is the BIC ARFIMA (0, d, 0) model that has a surprisingly high rank order. Inspection of the weights confirmed that for 1 participant, the ARFIMA (0, d, 0) model performed particularly well, whereas it performed relatively poorly for the majority of participants.

For the Van Orden et al. (2003) simple RT experiment, then, Table 1 shows that no strong conclusions can be drawn with respect to the presence of persistent serial correlations or $1/f^{\alpha}$ noise: The AIC weights are relatively evenly distributed over transient ARMA and persistent ARFIMA models, and the BIC weight is highest for the transient ARMA (1, 0) model.

Table 2 shows the results of the same analyses applied to the data from Van Orden et al.'s (2003) Experiment 2, the word-naming task. Again, AIC weights are spread out over a relatively wide range of models, whereas BIC weights concentrate mostly at the models that have only few free parameters. In addition, AIC prefers the ARFIMA (2, d, 1) model, as it did for the simple RT experiment. The ARFIMA (2, d, 1) model also has the best AIC value for 7 out of 20 participants. Several other ARFIMA models also perform well, and the evidence for persistent correlations is somewhat stronger here than in the simple RT experiment. Nonetheless, the summed AIC weight for all ARMA models is a respectable .34 and cannot be ignored. According to the BIC, the decision is between ARMA (1, 0) and ARFIMA (0, d, 0), again as in the simple RT experiment. This time, the ARFIMA (0, d, 0) model scores a clear win, $w_{0,d,0}(\mathrm{BIC}) = .47$, even though the ARMA (1, 0) model cannot be ignored, as its weight, $w_{1,0}(\mathrm{BIC}) = .21$, is almost half of that for the ARFIMA (0, d, 0) model. For the Van Orden et al. naming experiment, then, persistent ARFIMA models tend to outperform transient ARMA models, although arguably not by a wide margin.

In sum, the model-fitting procedures showed some support for the existence of persistent serial correlations. However, this support does not appear to be very strong. We conclude, therefore, that the transient ARMA models are competitive and cannot be excluded from consideration on the basis of the results from Van Orden et al. (2003).

Table 2

Average Model Weights, Average Rank Order, and Number of Rank Order Ones (i.e., Wins) for ARMA and ARFIMA Models Fitted to Van Orden, Holden, and Turvey's (2003) Experiment 2 (i.e., Word Naming), Separately for AIC and BIC

	AIC				BIC			
	Weight				Weight			
Model	*M*	*SD*	Rank	Wins	*M*	*SD*	Rank	Wins
Transient correlation								
ARMA (0, 0)	.00	.00	18.0	0	.00	.00	16.8	0
ARMA (1, 0)	.04	.06	10.9	**2**	**.21**	**.24**	5.4	**6**
ARMA (2, 0)	.04	.05	8.9	0	.03	.05	7.7	0
ARMA (0, 1)	.02	.04	14.3	0	.09	.13	8.9	0
ARMA (0, 2)	.03	.06	11.4	**2**	.03	.10	9.5	1
ARMA (1, 1)	.06	.06	7.3	1	.04	.08	**5.1**	0
ARMA (2, 1)	**.06**	**.05**	7.3	0	.00	.00	9.3	0
ARMA (1, 2)	.06	.04	**7.1**	1	.00	.00	8.8	0
ARMA (2, 2)	.03	.02	11.3	0	.00	.00	14.9	0
Persistent correlation								
ARFIMA (0, *d*, 0)	.07	.08	8.0	3	**.47**	**.35**	**2.5**	**12**
ARFIMA (1, *d*, 0)	.06	.03	**6.4**	0	.03	.02	4.5	0
ARFIMA (2, *d*, 0)	.03	.02	10.3	0	.00	.00	10.9	0
ARFIMA (0, *d*, 1)	.05	.03	7.1	0	.03	.02	5.2	0
ARFIMA (0, *d*, 2)	.04	.04	10.8	0	.00	.01	11.0	0
ARFIMA (1, *d*, 1)	.06	.08	9.1	1	.04	.18	9.4	1
ARFIMA (2, *d*, 1)	**.14**	**.14**	7.0	**7**	.00	.01	12.3	0
ARFIMA (1, *d*, 2)	.09	.12	9.3	1	.00	.02	13.2	0
ARFIMA (2, *d*, 2)	.11	.15	6.9	2	.00	.00	16.1	0

Note. *Weight* denotes the average model weights (i.e., averaged over participants). *Rank* denotes the average rank order of the models (i.e., averaged over participants). The ARMA and ARFIMA models that are best in terms of average weight, average rank order, and total number of wins are in bold. AIC = Akaike's information criterion; BIC = Bayesian information criterion; ARMA = autoregressive moving average; ARFIMA = autoregressive fractionally integrated moving average.

Does Human Cognition Really Behave Like a Pile of Sand?

In this section, we consider more conceptual issues and focus on what the theoretical implications would be if the presence of $1/f^{\alpha}$ noise had been indisputable. How much evidence would such a finding yield for the hypothesis that the workings of human cognition are similar to the behavior of a pile of sand?

Certainly, the prediction from a SOC account of $1/f^{\alpha}$ noise in human cognition is, at first sight, very surprising. We agree with Roberts and Pashler (2000) that the confirmation of surprising predictions constitutes strong support for a theory. SOC also distinguishes itself by addressing the trial-to-trial temporal dynamics of human behavior, in contrast to many current models for information processing in simple laboratory tasks that do not address such issues (but see Botvinick, Braver, Barch, Carter, & Cohen, 2001, pp. 640–643; Laming, 1968). However, several arguments suggest that even if Van Orden et al. (2003) had convincingly demonstrated the presence of $1/f^{\alpha}$ noise in their data, such a demonstration would not have provided the strong support for the SOC hypothesis that Van Orden et al. seem to have implied.

First, the finding of $1/f^{\alpha}$ noise in human cognition is not universal, despite Van Orden et al.'s (2003) claims to the contrary. In fact, some experiments find no serial correlations at all (e.g., Busey & Townsend, 2001; Townsend, Hu, & Kadlec, 1988), whereas other experiments find only exponentially decaying, transient correlations (e.g., Laming, 1968, 1979). When one associates, as did Van Orden et al., the presence of $1/f^{\alpha}$ noise with the presence of SOC, the finding of serial dependencies other than those of the $1/f^{\alpha}$ type is problematic; such a finding casts doubt on the robustness and generality of the claim that the human brain operates as a SOC system. In response, proponents of the SOC hypothesis may argue that the human brain operates as a SOC system only under certain specific experimental conditions. It may also be argued that external and perhaps unidentified experimental factors can obscure the $1/f^{\alpha}$ noise process or can turn it into a transient ARMA process. However, these arguments certainly require further explanation (i.e., how exactly would a $1/f^{\alpha}$ noise process be transformed into an ARMA process?), and in the absence of a process model that addresses the origin of serial correlations, the above arguments appear to be ad hoc.

Second, even if all SOC systems necessarily gave rise to $1/f^{\alpha}$ noise, the presence of $1/f^{\alpha}$ noise does not guarantee that the generating system is a SOC system. Jensen (1998) 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" (p. 13).[4] One such explanation is popular in econometrics and is due to Granger (1980). Granger hypothesized that the $1/f^{\alpha}$ noise observed in global economic measures comes about via aggregation of multiple component processes that separately generate transient correlations. Specifically, assume each of a possible infinite number of component processes, $X_t(i)$, $i = 1, 2, \ldots$, is a simple independent first-order autoregressive process, that is, AR(1), where behavior at time t depends partially on behavior at time $t-1$: $X_t(i) = \phi_i X_{t-1}(i) + \varepsilon_t(i)$, where $\phi_i \in (-1,1)$ and ε_t is an independent white-noise process with mean zero and variance σ_i^2. It is well-known that each of these component AR(1) processes generates transient rapidly decaying serial correlations. The overall behavior of the system is assumed to be a simple aggregation of the behavior of the component processes: $X_t = \sum_{i=1}^{\infty} X_t(i)$. When the parameters ϕ_i come from a beta distribution with suitable parameters, Granger (1980) showed that the aggregate series X_t displays persistent serial correlations (see also Beran, 1994, pp. 14–16). Granger's account can easily be applied to human cognition. All that needs to be assumed is that the observed behavior is an aggregation of the behavior of many independent groups of neurons, each with their own different autoregressive decay parameter (cf. Chen, Ding, & Kelso, 2001; Ding, Chen, & Kelso, 2002).

Several other models are able to mimic $1/f^{\alpha}$ noise for series lengths that are used in the study of human cognition (cf. Wagenmakers et al., 2004). Among other explanations, we suggested a regime-switching model for $1/f^{\alpha}$ noise in human cognition. In a regime-switching model, performance jumps discretely from one level to another, such as when the participant suddenly adopts a different response criterion for some number of trials or suddenly switches to a different strategy for performing the experimental task. Van Orden et al. (2003) asserted that the existence of different levels of performance in regime-switching models will give rise to "terraces" (p. 344) in performance and that these terraces are never observed in human performance. In practice, we have observed that when a simple AR(1) process is added to the regime-switching process, the noise in the resulting time series makes it difficult to detect terraces in simulated data sets. In addition, we note that one of the characteristic qualitative features of $1/f^{\alpha}$ noise is that "there are relatively long periods where the observations tend to stay at a high level, and on the other hand, there are long periods with low levels" (Beran, 1994, p. 41).

Van Orden et al. (2003) focused on how RT fluctuates from trial to trial. As Van Orden et al. pointed out, sequential sampling models such as the diffusion model (i.e., a continuous-time random walk model; cf. Ratcliff, 1978) are currently very successful in describing several aspects of RT performance (e.g., RT distributions for correct responses and error responses, the effects of speed–accuracy manipulations, the effects of aging; see, e.g., Ratcliff, Thapar, & McKoon, 2001; Ratcliff, Van Zandt, & McKoon, 1999). The diffusion model incorporates trial-to-trial fluctuations (e.g., Ratcliff & Rouder, 1998, 2000; Ratcliff & Smith, 2004; Ratcliff & Tuerlinckx, 2002) in various components of processing such as drift rate and starting point. However, this across-trial variability is random, and, until recently, the diffusion model had not been applied to long-range dependencies or $1/f^{\alpha}$ noise (for an application to short-range sequential effects, see Ratcliff, 1985; Ratcliff et al., 1999).

In another article (Wagenmakers et al., 2004), we proposed that sequential sampling models such as the diffusion model could be extended to handle the observed serial correlations in two-choice tasks by incorporating changes on different time scales. This proposal is in agreement with Treisman and Williams (1984), who developed a model for sequential effects in which criterion setting involves two processes: a short-term process that adjusts the criterion on a trial-by-trial basis (i.e., based on the sequence of prior decisions) and a relatively long-term process that is based on

[4] Another explicit demonstration that power laws are not uniquely associated with SOC is provided by Newman (1997).

factors such as prior knowledge or payoff values (Treisman & Williams, 1984, pp. 93–97).

Van Orden et al.'s (2003) main objection against sequential sampling models such as the diffusion model was that the model is "one part of a largely unspecified system" (p. 334). In reply, we point out that the generality of the diffusion model framework allows it to be successfully applied to a wide array of experiments from different paradigms. In the diffusion model, various sensory input mechanisms transmit information into decision stages that are related to various output mechanisms (vocal, manual, saccadic). This level of generality is needed if the diffusion model is to account for two-choice decisions for a variety of stimulus and cognitive dimensions. Also, the predictions of the diffusion model can be rigorously derived by analytical methods or simulations, and in this sense it is very specific—this is in sharp contrast to the framework of SOC as currently proposed by Van Orden et al. Further, recent work has used single-cell recordings in the monkey saccade system in an attempt to ground the diffusion model more firmly in biological reality (e.g., Ratcliff, Segraves, & Cherian, 2003). In simple two-choice tasks, the behavior of the cells under study (movement-sensitive cells in the frontal eye field, movement-sensitive cells in the lateral interparietal cortex, and buildup/prelude cells in the superior colliculus) appears to involve the accumulation of evidence toward one of two decisions, as predicted by the diffusion model. This conclusion is based on the fact that the aggregate behavior of neuronal activity follows the time course of accumulation of information within the decision model fit to the behavioral data (Gold & Shadlen, 2001; Ratcliff et al., 2003; Roitman & Shadlen, 2002; Smith & Ratcliff, 2004).

Third, not all SOC systems generate $1/f^{\alpha}$ noise under all conditions. Typically, SOC systems only generate $1/f^{\alpha}$ noise for specific dependent variables under certain specific conditions. For instance, in the case of the self-organizing pile of sand, the total mass of the sand pile shows $1/f$ noise across a wide range of frequencies (for details, see Jensen, 1998, pp. 30–42). However, this only happens when the new grains of sand are added along the edges (i.e., along the two orthogonal walls)! When the grains of sand are added to random positions on the interior of the pile, there is no $1/f$ noise (see Jensen, 1998, p. 42). Further, certain specific piles do not generate $1/f$ noise when they are made up of grains of sand, but they do generate $1/f$ noise when they are made up of grains of rice (for details, see Jensen, 1998). Because the nature of the interacting primitives in an SOC theory of human cognition is not clear, it is also not clear that SOC will necessarily lead to $1/f$ noise in human cognition. Also, Van Orden et al. (2003) did not specify how an SOC system can go from a perfect intrinsic $1/f$ noise pattern to the observed pattern of $1/f^{\alpha}$, with α much smaller than 1. Van Orden et al. mentioned that the intrinsic pattern is "decorrelated" (p. 338) by external factors, but how adding noise to a $1/f$ process reduces α in SOC systems is not specifically explained. It might well be possible to make a SOC system produce $1/f^{\alpha}$ with α much smaller than 1, but this has not been shown by the authors in any convincing way (i.e., via simulations or analytic methods).

Fourth, it is not clear what kind of new predictions follow from the characterization of human cognition as a pile of sand other than that it predicts persistent serial correlations (under certain situations and for certain dependent variables). How can the SOC framework be applied to human cognition except to note the persistence of serial correlations? Although the components, variables, or parameters in many dynamical systems models have some clear psychological meaning, the nature of the interacting primitives and the manner in which they give rise to the spectrum of complex human behaviors is not apparent from Van Orden et al.'s (2003) proposal.

A final, related argument is that the account, as it currently stands, is underspecified. When SOC systems have been proposed for sand piles, earthquakes, forest fires, evolution, or populations of neurons (da Silva, Papa, & de Souza, 1998; Usher, Stemmler, & Olami, 1995), such systems have always been implemented as specific models. This makes it possible to identify precisely what makes the models behave as they do. Also, such models might then be subjected to additional tests. We found it very difficult to extract from the Van Orden et al. (2003) article how exactly to apply the SOC framework to human cognition other than in very general terms. We hazard to guess that the authors' view is that the slow driving of SOC systems (cf. adding grains of sand to the pile) corresponds to the gradual accrual of information and that when the threshold is passed, the system responds (cf. avalanches in the sand pile). After the response is made, the state of the system is relaxed, and the process of information accrual starts again when a new stimulus is presented.

The above framework could be an interesting start, but to apply it to human cognition, several details need to be specified. The fact that information gradually accumulates until a decision threshold is passed is inherent in the majority of psychological models of choice response time (cf. Ratcliff & Smith, 2004). These models, such as the diffusion model (e.g., Ratcliff, 1978; Stone & Van Orden, 1993), might be underspecified in certain areas, as Van Orden et al. (2003) argued they are, but at least they are specified in enough detail to allow a wide range of data to be successfully described and, more important for scientific rigor, predicted. We believe the SOC framework proposed by Van Orden et al. is intellectually stimulating and interesting, but we challenge the authors to develop a specific model for how SOC operates in human cognition. It could be, for instance, that such a specific model would produce time series that are uncharacteristic of human performance, or it could be that correct behavior of such a model would depend on crucial assumptions that can be empirically tested.

Summary

The analyses reported by Van Orden et al. (2003) do not support their claim of $1/f^{\alpha}$ noise in repeated measurements of human performance or the implication of SOC. We argued that the evidence in favor of persistent serial correlations should be evaluated against the evidence for transient serial correlations, and we showed how this may be accomplished using the family of ARFIMA time series models. We believe that the finding of $1/f^{\alpha}$ noise constitutes only circumstantial evidence for the framework of SOC. The SOC framework does not necessarily lead to $1/f^{\alpha}$ noise under all circumstances or with respect to all dependent variables (Jensen, 1998). This observation underscores the need for a specific implemented model for SOC in human cognition. We briefly presented a simple alternative model (Granger, 1980) that can account for $1/f^{\alpha}$ noise by the aggregation of many independent autoregressive processes that together determine behavior. The above considerations strongly suggest that the paradigm shift proposed by Van Orden et al. is premature and that such an approach requires the application of specific, empirically testable, and preferably quantitative models of human behavior.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716–723.

Alstrøm, P., & Stassinopoulos, D. (1995). Versatility and adaptive performance. *Physical Review E, 51,* 5027–5032.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics, 73,* 5–59.

Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York: Springer-Verlag.

Bak, P., & Chialvo, D. R. (2001). Adaptive learning by extremal dynamics and negative feedback [Electronic version]. *Physical Review E, 63,* 031912.

Bak, P., & Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters, 71,* 4083–4086.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of 1/*f* noise. *Physical Review Letters, 59,* 381–384.

Beran, J. (1994). *Statistics for long-memory processes*. New York: Chapman & Hall.

Bisaglia, L. (2002). Model selection for long-memory models. *Quaderni di Statistica, 4,* 33–49.

Bisaglia, L., & Guégan, D. (1998). A comparison of techniques of estimation in long-memory processes. *Computational Statistics & Data Analysis, 27,* 61–81.

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.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Busey, T. A., & Townsend, J. T. (2001). Independent sampling vs. interitem 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. (2001). Origins of timing errors in human sensorimotor coordination. *Journal of Motor Behavior, 33,* 3–8.

Chialvo, D. R., & Bak, P. (1999). Learning from mistakes. *Neuroscience, 90,* 1137–1148.

da Silva, L., Papa, A. R. R., & de Souza, A. M. C. (1998). Criticality in a simple model for brain functioning. *Physics Letters A, 242,* 343–348.

Davidsen, J., & Lüthje, N. (2001). Comment on "1/*f* noise in the Bak-Sneppen model" [Electronic version]. *Physical Review E, 63,* 063101.

Davidsen, J., & Paczuski, M. (2002). $1/f^{\alpha}$ noise from correlations between avalanches in self-organized criticality [Electronic version]. *Physical Review E, 66,* 050101.

Davidsen, J., & Schuster, H. G. (2000). $1/f^{\alpha}$ noise from self-organized critical models with uniform driving. *Physical Review E, 62,* 6111–6115.

Davidsen, J., & Schuster, H. G. (2002). Simple model for $1/f^{\alpha}$ noise [Electronic version]. *Physical Review E, 65,* 026120.

Ding, M., Chen, Y., & Kelso, J. A. S. (2002). Statistical analysis of timing errors. *Brain and Cognition, 48,* 98–106.

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, 41,* 333–348.

Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2004). *ARFIMA time series modeling of serial correlations in human performance*. Manuscript submitted for publication.

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, March 24). 1/*f* noise in human cognition. *Science, 267,* 1837–1839.

Gisiger, T. (2001). Scale invariance in biology: Coincidence or footprint of a universal mechanism? *Biological Reviews of the Cambridge Philosophical Society, 76,* 161–209.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science, 5,* 10–16.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics, 14,* 227–238.

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.

Jensen, H. J. (1998). *Self-organized criticality.* Cambridge, England: 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.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.

Laming, D. R. J. (1979). Autocorrelation of choice-reaction times. *Acta Psychologica, 43,* 381–412.

Linkenkaer-Hansen, K. (2002). *Self-organized criticality and stochastic resonance in the human brain.* Unpublished doctoral dissertation. Retrieved May 21, 2003, from http://lib.hut.fi/Diss/2002/isbn9512262177/isbn9512262177.pdf

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., & Ilmoniemi, R. J. (2001). Long-range temporal correlations and scaling behavior in human brain oscillations. *The Journal of Neuroscience, 21,* 1370–1377.

Malamud, B. D., Morein, G., & Turcotte, D. L. (1998, September 18). Forest fires: An example of self-organized critical behavior. *Science, 281,* 1840–1842.

Myung, I. J., Forster, M. R., & Browne, M. W. (Eds.). (2000). Model selection [Special issue]. *Journal of Mathematical Psychology, 44*(1–2).

Newman, M. E. J. (1997). Evidence for self-organized criticality in evolution. *Physica D, 107,* 293–296.

Ooms, M., & Doornik, J. A. (1999). *Inference and forecasting for fractional autoregressive integrated moving average models with an application to US and UK inflation* (Econometric Institute Report EI-9947/A). Rotterdam, the Netherlands: Erasmus University, Econometric Institute.

Paczuski, M., Maslov, S., & Bak, P. (1996). Avalanche dynamics in evolution, growth, and depinning models. *Physical Review E, 53,* 414–443.

Priestley, M. B. (1981). *Spectral analysis and time series*. London: Academic Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163.

Rangarajan, G., & Ding, M. (2000). Integrated approach to the assessment of long range correlation in time series data. *Physical Review E, 61,* 4991–5001.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review, 92,* 212–225.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347–356.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 127–140.

Ratcliff, R., Segraves, M., & Cherian, A. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. *Journal of Neurophysiology, 90,* 1392–1407.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111,* 333–367.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16,* 323–341.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9,* 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, R. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 102,* 261–300.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.

Roitman, J. D., & Shadlen, M. N. (2002). Responses of neurons in the lateral interparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22,* 9475–9489.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Slifkin, A. B., & Newell, K. M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science, 7,* 170–177.

Smith, P. L., & Ratcliff, R. (2004). The psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27,* 161–168.

Sornette, D. (2000). *Critical phenomena in natural sciences.* Berlin, Germany: Springer.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744–774.

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.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68–111.

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

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition. *Psychonomic Bulletin & Review, 11,* 579–615.

Ward, L. M. (2002). *Dynamical cognitive science.* Cambridge, MA: MIT Press.

Received July 1, 2003
Revision received February 27, 2004
Accepted March 1, 2004 ■