Sequence Effects in Estimation

David J. Elliott – B.Sc, B.Psyc. (Hons.)

Thesis submitted for Master of Science.

September 2006, University of Newcastle, Australia

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other university of institution.

Signed:

Acknowledgments & Dedications

I would like to thank everyone who has contributed to this thesis in one way or another.

Particularly, I would like to thank my supervisors, Andrew Heathcote ad Rachel Heath, who are always in a rush, but always stop to provide patient help, direction, motivation, and inspiration.

Big thanks to all the people who have come and gone through the cognitive science laboratory, for your company and help. (See, I was actually working).

Thanks to my family and friends, here and gone, for their support and inspiration.

Endless thanks to all the anonymous people who endured these experiments.

This thesis is dedicated to nana and pop, Alec and Ella Wilson.

Wish you were here.

Table of Contents

Abstract

The apparently random fluctuations in serial interval estimation are examined for evidence of sequential structure. Specifically, fluctuations are examined for evidence of two phenomena: nonlinear dynamics, and 1/*f* noise. Both phenomena have been reported in cognitive performance data, and have also been reported in a wide range of natural and artificial systems in many fields. Three experiments examined a possible relationship between these two phenomena and task demands in interval estimation. Traditional and contemporary methods of identifying 1/*f* noise are discussed, evaluated and applied, and contemporary methods of identifying nonlinear dynamics are discussed and applied. Neither nonlinear dynamics nor 1/*f* noise are shown to be present in the interval estimation data. An alternative explanation for the appearance (and absence) of 1/*f* noise in cognitive performance is discussed.

Sequence Effects in Interval Estimation

The traditional approach to experimental psychology involves the search for a more definitive catalogue of human behaviour, in the manner of Donders (1868): A simple reaction takes *x* seconds, a reaction involving a choice takes *y* seconds, reactions involving *n* choices takes *z* seconds for each choice, etc. From this perspective, research involves the meticulous delineation of human behaviour by observing behavioural performance under systematically altered conditions, and testing for differences between conditions using inferential statistics such as Analysis of Variance (ANOVA).

However, no single observation provides a definitive measure of behaviour. No device measures reaction time infallibly, no two people react at precisely the same speed, no person responds twice at precisely the same speed, and no measurement can utterly escape the influence of the world beyond the boundaries of the experiment (Kaplan & Glass, 1995; Luce, 1995; Thelen & Smith, 1994). To resolve this issue, experimental psychology adopted a statistical concept from the physical sciences: the "error law," based on the Gaussian or Normal distribution, or "bell curve" (Stewart, 1989). According to the error law, observed values of natural phenomena cluster around their true value as independent samples drawn from a Gaussian probability distribution. When a sufficient number of repeated measurements are averaged, the average provides the most reliable estimate of the true value. When repeated measures taken from the same device are averaged, measurement error is minimised; when repeated measures taken from different individuals are averaged, individual differences are minimised; and when repeated measures from the same individual are averaged, individual error is minimised. Conventional analysis, therefore, averages repeated measures, and measures between individuals, in an attempt to produce a single, reliable estimate of the ‘true’ measure of behaviour observed under a specific condition (Figure 1), which can then be compared to averaged measures of other observed behaviours using inferential statistics such as analysis of variance (ANOVA).

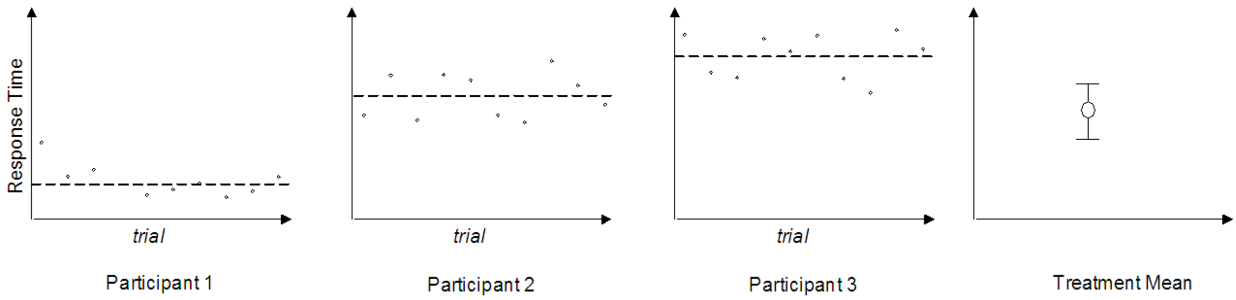


Figure 1. A simplified illustration of the traditional experimental approach, showing repeated measurements taken from participants 1, 2, 3 in a single treatment condition. Individual scores are averaged (dashed lines) to remove trial-by-trial variation. Means are then averaged across subjects to remove individual differences to produce a single score, the treatment mean (right, with standard error bar).

For example, Chen, Repp and Patel (2002) studied rhythmic finger tapping at a tempo of 500ms using auditory, visual and combined audio-visual metronomes. Six participants performed approximately 300 trials for each metronome condition. Asynchronies (the latency between metronome signal and tap) were examined by averaging across trials and subjects in each condition, yielding a mean and standard error for each condition (Figure 2). Inferential testing compared treatment means using

Analysis of Variance (ANOVA), which revealed significantly smaller asynchronies (i.e. more accurate synchrony) using the visual metronome in comparison to the auditory and combined audio-visual metronomes.

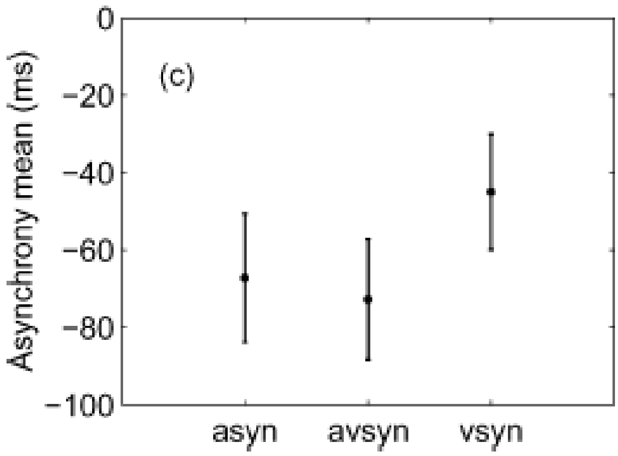


Figure 2. Treatment means and standard error bars of asynchronies for auditory (asyn), audio-visual (avsyn) and visual (vsyn) metronomes (reproduced from Chen, Repp & Patel, 2002).

This comparison exemplifies the conventional static approach to analysis, the foundation of classical experimental psychology. However, in neglecting the residual error, or noise, this approach ignores two potentially valuable sources of information: individual differences, and inter-trial variation. Gilden (1997) examined variance in reaction times for mental rotation, lexical decision, serial visual search, and parallel visual search, and concluded that variation in treatment means accounted for only a small proportion of variation in response times. Variation due to treatments in a mental rotation task, for example, accounted for only 10% of the total variation in reaction times. Heath (2000) warned: “an inconceivable amount of deterministic evidence from over a century of published psychological research might be concealed in the error terms of ANOVA models." Even if an exaggeration, Heath's warning nonetheless highlights the potential cost of neglecting these sources of information.

Neglecting inter-trial variation may also invalidate a core assumption of the conventional statistical analysis techniques. Since statistical methods based on the General Linear Model, such as Analysis of Variance, attribute residual error to the 'natural error law,' they assume that the residual error can be characterized as independent identically distributed (i.i.d) random variables. Yet in a number of experimental tasks evidence suggests that residual error is not independent, but instead shows sequence effects (e.g. Bertelson, 1961; Brewer & Smith, 1984; Busey & Townsend, 2001; Kornblum, 1973; Lamming, 1968; Rabbit, 1992; Townsend, Hu & Kadlec, 1988; Wertheimer, 1953). Therefore, rather than assume independence, it becomes necessary to distinguish independence from dependence. Serial dependence, or sequence effects, in data indicate that one or more previous trials influence the current trial to some degree. That is, as a result of the previous history of trials and responses, an imprint (in some form) is made, which affects responses in the future (Fecteau & Munoz, 2003). The series in Figure 1 demonstrate sequence effects in the form of simple periodicities: the response times for participant 1, for example, increase slightly on each successive trial, then drop periodically. Such periodic behaviour is inconsistent with the assumption of independent trials, making conventional analytic methods such as ANOVA at best approximate, and at worst inappropriate. Yet, with appropriate analysis methods, such inter-trial variance may itself provide insight into the processes involved in response times. The present study focuses on sequence effects in response time tasks, with the aim of exploiting these often overlooked potential sources of information, to gain greater insight into performance in response time tasks.

Sequence Effects in Experimental Psychology

One of the simplest examples of a sequence effect occurs when guessing the outcome of a coin toss. The outcome of a fair coin toss is independent of the previous history of coin tosses. That is, the outcome of any toss is equally likely to be heads or tails, regardless of whether the previous outcome was heads or tails. However, Jarvik (1951) found that people tend to guess “tails” when their previous guess of “heads” was correct. Jarvik called this a “negative recency effect,” which he attributed to poor understanding of probability. Specifically, Jarvik attributed this effect to the Gambler’s fallacy: the mistaken belief that the probability of an event with fixed probability changes as a result of previous outcomes.

The Gambler’s fallacy extends well beyond coin tossing. An earlier study by Fernberger (1920) demonstrated the gambler’s fallacy affecting responses in a weight discrimination task. Fernberger presented participants with a series of brass weights and asked the participants whether the weight was “lighter,” “equal,” or “heavier” than a comparison brass weight. Participants tended to respond with “heavier” if their previous response had been “lighter,” and respond with “lighter” if their previous response had been “heavier.” Fernberger called this a “contrast effect.”

Arons and Irwin (1932) replicated Fernberger’s study, but used equal brass weights on each trial. They found the same contrast effect in successive responses, despite using equal weights. They concluded that the ‘contrast effect’ of responses, or gambler's fallacy, was independent of the stimuli presented. Preston (1936) showed that the contrast effect extended back to the previous two trials: participants not only avoided repeating the response to the previous trial, but avoided repeating the response on the penultimate trial. For example, if the participant responded with "lighter" on their first trial, they might respond with "equal" on the second trial, and respond with "heavier" on the third trial.

Sequence effects appear to be innate phenomena. A study by Canfield and Haith (1991) observed sequence effects in behaviour of 2-3 month old infants. When presented with visual stimuli that alternated between the left and ride sides of a display, two-month old infants would fixate on the next position after only two minutes of exposure to the task. Three month old infants could anticipate asymmetric sequence patterns (e.g. left-left-right, right-right-left). This suggests an innate, rapidly developed capacity for understanding and utilising sequence effects to facilitate prediction of future events in a sequence.

Sequence Effects in Response Times (RT)

Hyman (1953) inadvertently provided one of the first reports of sequential effects in response time tasks. Hyman presented participants with two concentric squares with a globe set in the corner of each square. He illuminated a single globe in a random location with an inter-stimulus interval (ISI) of approximately 10 seconds. Participants identified the location of the signal vocally. Hyman noted in his results that responses were unusually fast when the same stimulus was presented on consecutive trials. That is, response time decreased with stimulus repetitions.

Interest in sequence effects in response times increased in the 1960s, following the publication of two seminal studies conducted by Bertelson and Laming. Bertelson (1961) examined sequence effects in a simple two-choice RT task. The task presented participants with two neon globes and two response keys mapped to the globes, with participants required to press the key corresponding to each globe when illuminated. The study involved three conditions: a REP condition, with 75% probability of a repeated stimulus; a RAND condition, with 50% probability of a repetition; an ALT condition, with 25% probability of a repeated stimulus. Bertelson reported that response times in the REP condition were significantly faster than for the ALT condition, and significantly faster than for the RAND condition, but there was no significant difference in response times between the ALT and RAND conditions. He concluded that there is a "facilitation favouring repetitions" (p.98), or what is now commonly referred to as the "(Bertelson) repetition effect": response times are faster for repeated signals than non-repeated signals. As Bertelson notes, this result is contrary to the Gambler's fallacy, since participants seem to expect repetitions rather than non-repetitions. However, it should be noted that these results were obtained only after prolonged practice: the study consisted of a practice session of 1000 trials of the RAND condition; three experimental sessions, with 1000 trials of one condition (randomly ordered) in each; and a second set of three experimental sessions, with 1000 trials of one condition (randomly ordered) in each. The repetition effect was only significant in the second set of sessions.

Laming (1968) reported a number of sequence effects following errors in performance in a choice RT task. On the trial following an error, the probability of making an error decreased significantly, while the response time increased significantly.

The increase in response time was greatest when the trial following an error required the same response that was incorrect on the previous trial. Laming reported that this increase in response time following an error would continue for up to three trials following an error. These results, together with the results of Bertelson (1961), led to a surge of interest in sequence effects in choice RT in particular, and in RT more generally. Transient dependence now represents a commonly accepted feature of performance in choice RT (Kornblum, 1973). The repetition effect, for example, appears to be ubiquitous in choice RT, having been identified in four-, five- and eight-choice tasks (Hale, 1969; Hoyle & Gohlson, 1968; Kirby, 1975; Kornblum, 1967, 1968, 1969; Leonard, Newman & Carpenter, 1966; Rabbitt, 1965, 1968; Remington, 1971; Schvaneveldt & Chase, 1969; Smith, 1968).

The assumed transient nature of sequence effects is highlighted by early classification of these effects into two classes: first-order sequence effects, effects that occur as a result of the immediately preceding stimulus-response pair, and higher-order effects, effects that occur as a result of stimulus-response pairs that occurred more than one trial into the past (Kornblum, 1973). However, in contrast to these transient dependencies, more recent findings suggest the existence of persistent, or long-range, dependence in psychological data series (e.g. Aks & Sprott, 2003; Chen, Ding & Kelso, 1997; Chen, Rapp & Patel, 2002; Gilden, 1997, 2001; Gilden, Thornton & Mallon, 1995; Wagenmakers et al. 2004; Yamada, 1996). As with the discovery of short-range dependence decades before, the discovery of long-range dependence has been partly responsible for a surge of interest in dynamical approaches to psychological time series. Much of the interest can be attributed to an increasing body of evidence for two related dynamical phenomena, which may not only provide explanations for existing sequence effects, but may also provide evidence for previously unknown structure in experimental data: nonlinear dynamics and 1/*f* noise.

The Dynamical Approach

Where the conventional, static approach averages across repeated observations to remove ‘nuisance' fluctuations from analysis, the dynamical approach embraces the fluctuations offered by repeated measurements. Where the static approach ignores time, the dynamical approach treats time as another variable, another source of information (Auyung, 2000). Formally, dynamical analysis examines the evolution of a system for evidence of relationships, or dependencies, between observations ordered sequentially in time. Relationships between observations are expressed formally as a set of mathematical functions that determine the evolutionary behaviour of the system (Peak & Frame, 1994). These functions can be expressed in general form as y = f(x), where x is the input (for example, the value of the preceding observation), and y is the output (for example, the value of the next observation), and f() is the unspecified (or possibly unknown) function or relationship between input and output. Mathematically, these functions can be classified as either linear or nonlinear.

For a linear dynamical system, the function $y = f(x)$ possesses two important properties: additivity and homogeneity. The property of additivity can be expressed algebraically as $f(x + y) = f(x) + f(y)$. The property of homogeneity can be expressed algebraically as $f(ax) = af(x)$, where a is a scalar. Any dynamical system which possesses these two properties is classified as linear. These two properties result in the principle of superposition, where the net response of multiple linear functions is also the sum of responses from the individual linear functions. In simple terms, a linear system can be understood as the sum of its parts: if the solution of $f(x)$ and $f(y)$ are known, then the solution of $f(x+y)$ is also known.

Stewart (1989) illustrates the principle of superposition with the example of ripples on a pond caused by throwing stones into the water. The equation that describes the motion of a shallow wave on a liquid surface is linear. When a stone is thrown into a pond, this linear equation provides an adequate approximation of how the surface ripples behave. When two stones are thrown into a pond, the solution of two wave equations, centred appropriately, describes how the two sets of ripples behave. In general terms, the solution of an n-stones problem is simply the sum of solutions for n one-stone problems.

However, dynamical models that approach the complexity of observed phenomena typically require nonlinear functions. A nonlinear function is classified as any function which does not possess the properties of additivity and homogeneity, and therefore does not obey the principle of superposition. A nonlinear system cannot be understood as the sum of its parts, which makes the solutions of nonlinear systems much more difficult, if not impossible.

An important distinction between linear and nonlinear functions is that if x and y are related by a linear function, a change in the value of x will result in a proportional change in y, and vice versa. However, if x and y are related by a nonlinear function, the proportion of change in the value of y depends on the value of x. A small change in the value of x may result in an extremely large change in the value of y, or an extremely large change in the value of *x* may result in only a small change in *y*. Gleick (1986) illustrates this distinction using the example of an ice-hockey puck. Ignoring the effect of friction, the acceleration of a puck on ice becomes a simple linear function of the force applied to the puck. However, when accounting for friction, the acceleration of the puck becomes a nonlinear function, because the friction acting on the puck depends on the speed of the puck.

To appreciate what makes the nonlinear dynamical approach attractive to experimental psychology - and increasingly to applied and clinical psychology (e.g., Gottman, Murray, Swanson, Tyson, & Swanson, 2005; Heiby, Pagano, Blaine, Nelson, Heath, 2003) – requires a more thorough explanation of some concepts of nonlinear dynamical theory.

In order to study a system, it is necessary to adequately specify the state of the system. A *state variable* specifies a property of the system necessary to describe the state of the system at any given time. In the bi-manual finger-tapping experiment of Haken, Kelso and Bunz (1985), the system could be adequately described using two state variables, one variable for the vertical position of each finger. The values of the two state variables – the vertical position of each finger - would adequately characterise the state of the system at any point in time. However, for complex systems involving many – possibly hundreds or thousands of state variables – it is necessary to characterise the state of the system using as few state variables as possible. The Haken et al. study, for example, reduced the two state variables to a single state variable: *relative phase*, defined as the vertical displacement of one finger relative to the other finger at time *t*. A single variable that adequately describes the global properties of the system, such as relative phase, is known as an order parameter.

Given one or more state variables, it is possible to visualise the evolution of the system. A simple method of visualizing its evolution would be to produce a time series plot: a plot of the state variable as a function of time. Figure 3 shows a time series plot of the *x* values of the three-dimensional (*x*, *y, z*) Lorenz differential equations:

$$\frac{dx}{dt} = \sigma(y - x)$$
$$\frac{dy}{dt} = x(r - z)$$
$$\frac{dz}{dt} = xy - bz$$

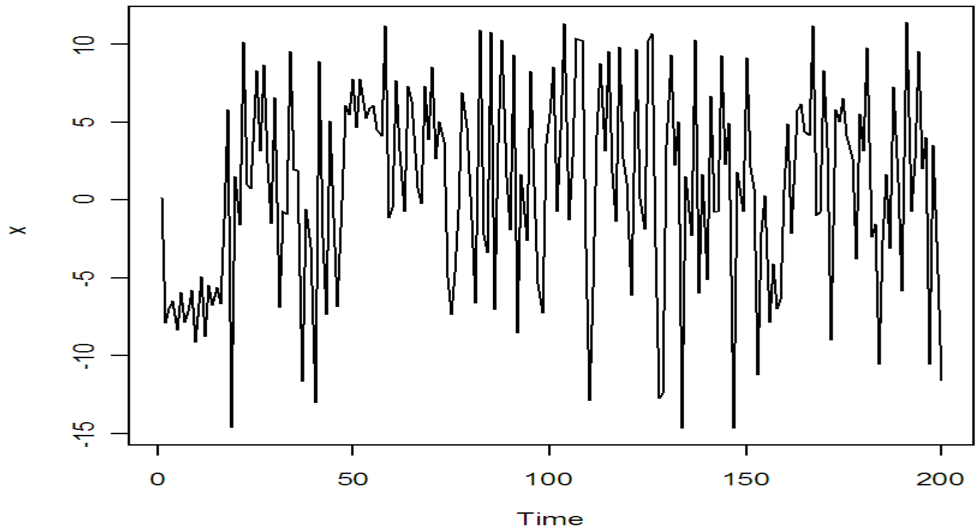


Figure 3. ***A time series plot of the first 200 x-values of the Lorenz differential equations.***

An alternate method of visualising the evolution of the system is a phase (or state) space plot: a plot of the relevant *n* state variables against each other in *n* dimensional space. On such a plot, a single point in phase space uniquely specifies a single state of the system (Figure 4). This allows the evolution of the system to be visualized by the evolution of the points in phase space, known as the trajectory of the system. A phase portrait is a plot of multiple potential trajectories of a system for different initial states.

The phase diagrams in Figure 4 show the evolution of the Lorenz differential equations in two-dimensions by plotting the *x* values as a function of the *y* values. Figure 4 shows the trajectory of the first 200 values (corresponding to the time series in Figure 3) as well as the trajectory of the first 5000 values, which produces the two-lobe Lorenz (or “butterfly”) attractor, the subset of phase space that “attracts” trajectories.

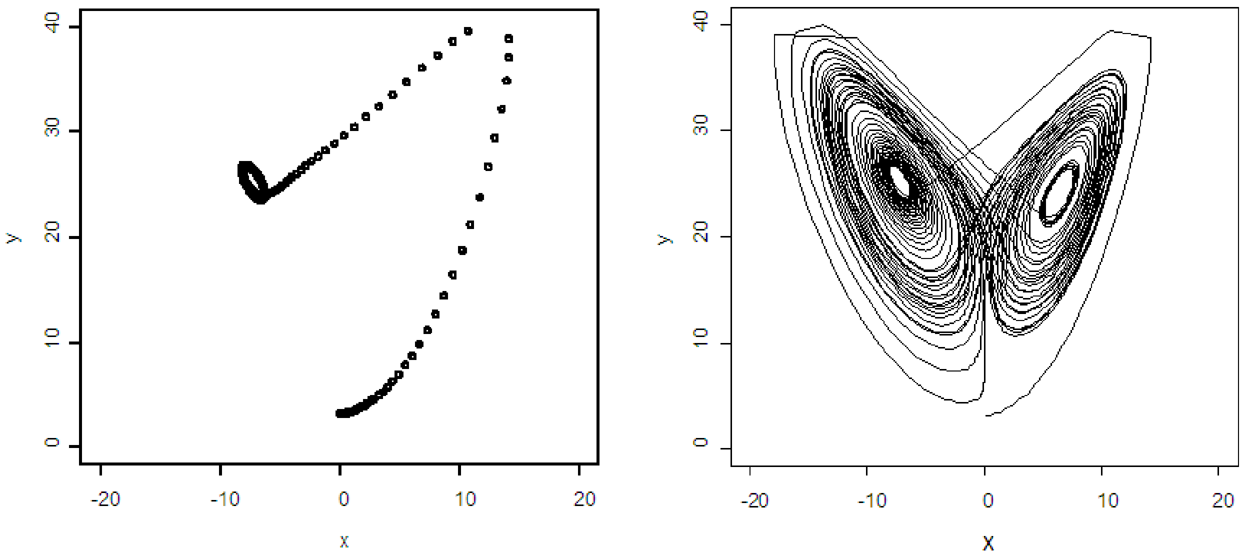


Figure 4. ***An example of a two-dimensional phase space diagram, the x and y values of the Lorenz differential equations, showing the trajectory of the first 200 values in the series (left) and the trajectory of the first 5000 values in the series plotted as a curve (right).***

Lastly, it may be necessary to define one or more system control parameters. A control parameter is a constant or variable that affects the behaviour of the system, but is controlled externally, analogous to an independent variable in conventional experimental methodology. In the Haken et al. (1985) bi-manual tapping experiment, the relevant control parameter is the frequency of tapping, controlled by altering the frequency of the metronome to which the participants must synchronise their tapping. By altering the frequency of the metronome, Haken et al. greatly affected the dynamics of the bi-manual finger tapping, as discussed in greater detail below.

The two-lobe Lorenz 'butterfly' attractor (Figure 4) demonstrates the trajectory of the system settling into a region of phase space. The behaviour of the system is limited to a trajectory within the vicinity of this region, with the trajectory 'attracted' to the attractor like iron filings are to a magnet. The attractor, therefore, explains a stable pattern of behaviour. For brevity, only three major classes of attractors will be discussed: point attractors, limit cycle attractors, and strange attractors.

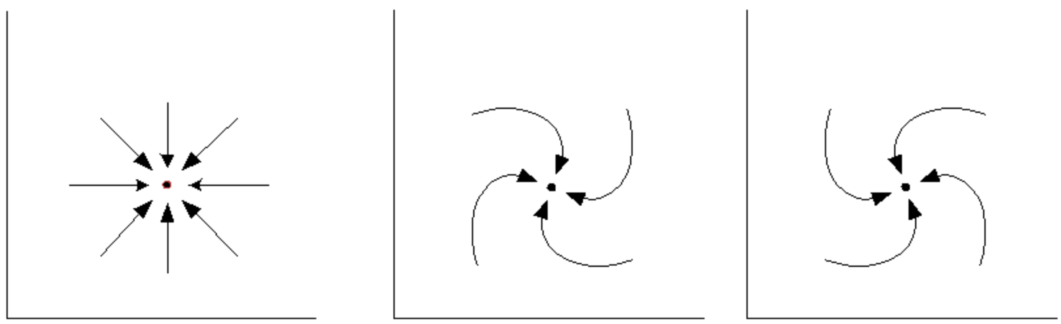


Figure 5. ***Examples of point attractors in phase space: a stable-node attractor (left), and stable-focus attractors approaching from different directions (middle and right).***

A point attractor, as the name implies, consists of a single point in phase space. If the system contains only a single point attractor, then all trajectories, regardless of initial states, are drawn to the point (Figure 5). A point attractor represents a system that reaches a steady state, where it no longer changes with time, including systems that inevitably come to rest. Trajectories can approach a point attractor either directly (stable-node attractor) or indirectly (stable-focus attractor).

A limit-cycle or periodic attractor consists of a closed loop in phase space. In a system containing a limit-cycle attractor, all trajectories converge toward a set of periodic steady states, with the system containing oscillating attractor values (Figure 6). A limit-cycle attractor represents a system where two or more states occur periodically. Note that trajectories inside and outside the cycle converge on to the cycle. As a result, a system already on the attractor can suffer perturbations and still return to its periodic state.

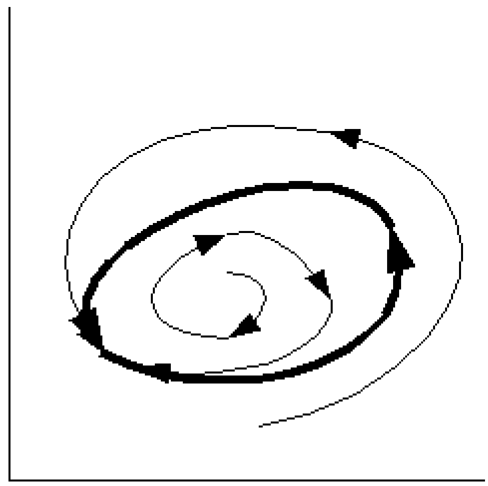


Figure 6. ***Example of a limit-cycle or periodic attractor (bold lines).***

More than two attractors can co-exist in the same phase space (Figure 7). When this occurs, the behaviour of the system depends on the initial state of the system, as the trajectory will converge on the attractor nearest to the initial state. One situation in which this occurs involves systems that undergo bifurcations, which will be discussed below.

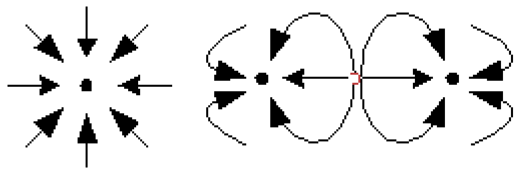


Figure 7. ***Examples of a point attractor (left) and two point attractors existing in the same phase space (right).***

In comparison to the simple geometry of the point and limit-cycle attractors, strange attractors have very complex geometrical structures. The two-lobe Lorenz ‘butterfly’ attractor (Figure 4) is one of the most well-known strange attractors. Strange attractors are a signature of a particular class of nonlinear systems known as ‘chaotic systems’ or ‘chaos,’ which possesses several unique characteristics. First, chaotic systems are non-periodic. Although the Lorenz attractor appears periodic, the trajectory does not pass through the same position in phase space more than once.

Second, chaotic systems are erratic yet deterministic. Although non-periodic to the point of appearing ‘random,’ chaotic systems are nonetheless predictable. Given the same initial conditions and the same parameter values, a chaotic system, with all its erratic, non-periodic behaviour, can be perfectly re-recreated, and future states can be perfectly predicted. However, in practice, this typically proves impossible due to the last characteristic of chaotic systems: sensitive dependence on initial conditions (Kaplan & Glass, 1995), known in popular culture as “the butterfly effect” (Gleick, 1986). Even an infinitesimal change in initial conditions can result in a disproportionate change in future behaviour. This property makes long-term prediction of chaotic systems, such as the weather, impossible. For example, the ‘butterfly effect’ states that the ability to predict the weather deteriorates rapidly because a prediction that fails to precisely define the current (initial) conditions – with a precision that includes the change in air pressure caused by a butterfly flapping its wings – will predict a weather system that rapidly diverges from the true weather system.

Changes between stable patterns of behaviour occur as a result of bifurcations. Bifurcations, in turn, occur as a result of changes in the control parameter. Figure 8 illustrates a change in behaviour of the logistic equation, a simple nonlinear system defined by:

$$x_{n+1} = rx_n(1 - x_n)$$

The change in behaviour results from an increase in the control parameter r. Figure 8 shows a diagram of a pitchfork bifurcation with corresponding 2-dimensional landscape models and phase portraits beneath. The 2-dimensional landscapes provide a metaphor for interpreting the effect of attractors. If the ball represents the state of the system at a point in time, the path of the ball rolling on the two-dimensional landscape under the influence of gravity characterises the trajectory of the system. The system (ball) is only stable when it settles at the base of the valley.

When the control parameter is sufficiently small, the logistic system exists in a stable state with a single point attractor (V1 in Figure 8). However, when the control parameter r reaches a critical value q, the system becomes unstable (V2 in Figure 8), and the previously stable state becomes unstable (represented by the white ball); however, two new potential stable states emerge (V3 & V4 in Figure 8). The system undergoes a transition from the unstable state to one of the new stable states. The result is a pitchfork bifurcation: a stable state becomes unstable, and a transition to bi-stability occurs. This allows change and flexibility, giving systems two options for change.

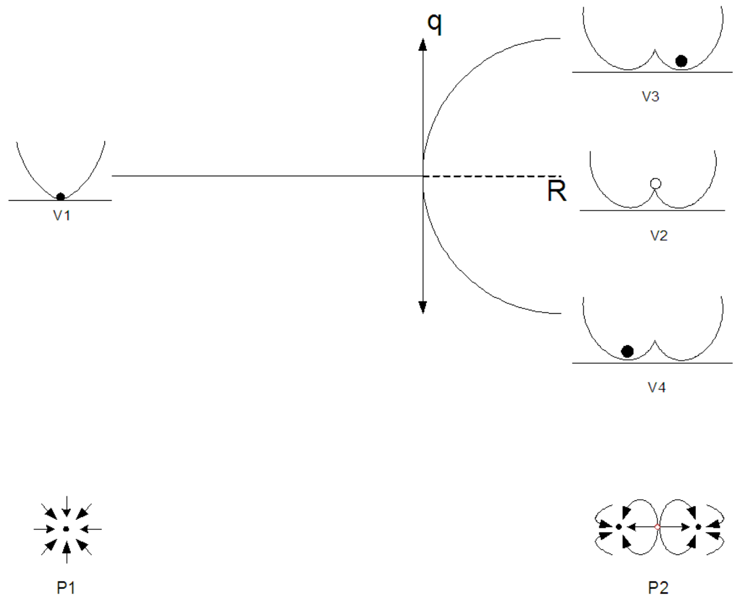


Figure 8. ***A bifurcation diagram overlaid with corresponding 'potential landscapes' (V1-V4), and phase portraits (P1-P2; adapted from Kelso, 1995).***

To illustrate the application of nonlinear dynamics, consider the Haken et al. (1985) bi-manual finger tapping study. Recall that the state variable consisted of the relative phase φ, in radians, of the two tapping index fingers, and the control parameter consisted of the finger tapping frequency, determined by the tempo of a metronome, with which the participants synchronised their tapping. Haken et al. had participants perform two tapping tasks: in-phase and anti-phase. In the in-phase task, participants had to tap their fingers at the same frequency and in the same direction; in the anti-phase task, participants had to tap their fingers at the same frequency and in opposite directions – that is, when one finger rose, the other finger fell. At low frequencies, participants performed both tasks with ease. However, when the frequency of the taps increased, participants had greater difficulty with anti-phase tapping (but not with in-phase tapping). At a critical frequency, the anti-phase tapping lapsed into in-phase tapping. Haken et al. explained these results in terms of an inverse bifurcation (Figure 9). When the frequency exceeds a critical value, the anti-phase tapping ($\varphi = +/- \pi$) becomes unstable (or repellers) and the in-phase tapping ($\varphi = 0$) becomes a stable point attractor.

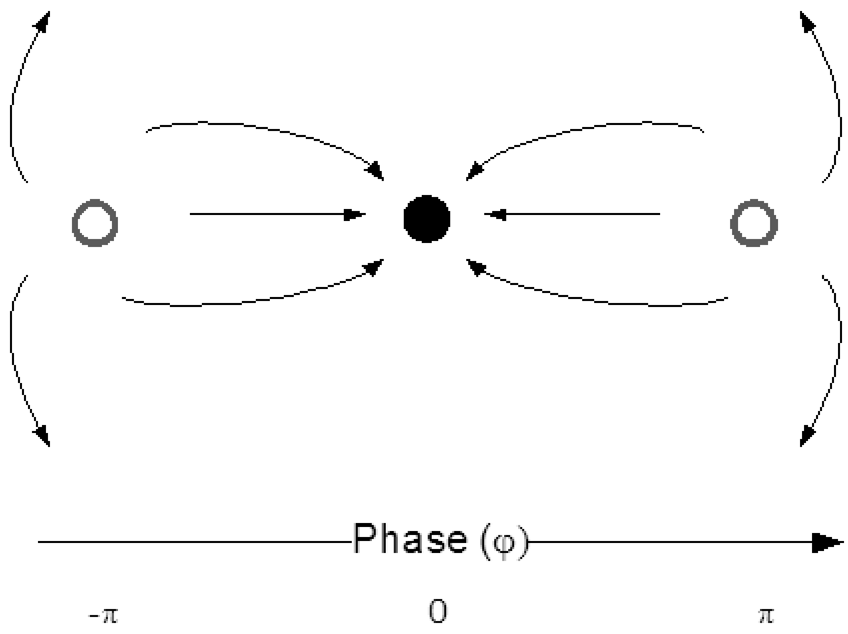


Figure 9. ***The Haken-Kelso-Bunz (HKB) model explained as an inverse bifurcation.***

The nonlinear dynamical approach provides a new framework for understanding human behaviour, by modelling behaviour as a system governed by nonlinear equations (e.g. Gottman et al., 2005; Thagard, 2005). According to this framework, attractors explain (in the formal sense) stable patterns of behaviour, while multiple attractors explain multiple stable patterns of behaviour. Abrupt changes in behaviour can be explained by phase transitions between attractors, while long term unpredictability of behaviour can be explained by sensitive dependence. Reflecting the vast array of possible stable behavioural patterns, attractors may take on a variety of forms: from the simple stable point attractors, through periodic attractors, to the erratic and unpredictable strange attractors of chaos (Tsonis, 1992).

1/f Noise

Another dynamical phenomenon receiving increasing attention in a variety of disciplines is 1/*f* noise. One of the most attractive features of 1/*f* noise is that it exhibits long-range dependence: the dependence between serial observations decays slowly with the number of intervening observations, so that the current observation depends to some extent on the value of observations far into the past. In formal terms, long-range dependence is defined in terms of the autocorrelation function, C(*k*), which indicates the degree of dependence, or correlation, between observations in a time series separated by an interval (lag) of *k* observations (Tsonis, 1992):

$$C(k) = \frac{Cov(x(t), x(t+k))}{\mathrm{var}(x(t))}$$

For a series of independent identically distributed (i.i.d) random variables, such as that assumed by the general linear model, C(k) = 0 for all lags *k*, indicating a complete lack of dependence between observations in the series. For a series exhibiting short-range dependence, such as the autoregressive model of RT proposed by Lamming (1968), C(k) decays rapidly towards zero with increasing k. For a series exhibiting long-range dependence, C(k) decays slowly towards zero. More precisely, for short-range dependence C(k) decays according to an exponential function, whereas for long-range dependence C(k) decays according to a power function (Beran, 1994).

Until recently, identification of 1/*f* noise in time series relied almost exclusively on power spectral density analysis, based on the Fourier transform of a time series. According to Fourier's law, any time series can be approximated by the sum of sine waves of varying frequencies, amplitudes and phases (Tsonis, 1992), where frequency (in Hz) is the number of cycles per second, amplitude is the magnitude (height) of the wave, and the phase is the point in the cycle where the first wave begins. The square of the amplitude provides the "power spectral density," which expresses the amount of variance (power) at each frequency. The power spectrum plots the power spectral density $S(f)$ for each frequency f, typically on log-log coordinates. In practice, spectral analysis typically relies on the Fast Fourier Transform, which provides a computationally efficient estimate of the Fourier Transform.

Power spectral density analysis relies on differential diagnosis, as different forms of noise possess distinctive power spectra. In the case of long-range dependence, the spectrum can be characterized by a power function, $S(f) = kf^{\alpha}$, where k is a constant, and α characterizes the class, or 'colour,' of the noise. In practice, identification of the class of noise relies upon an estimate of α provided by the slope of a linear regression of a bi-logarithmic $S(f)$ versus f plot. In simple terms, the degree of persistence of serial correlations relates to the amount of power observed at low frequencies: the proportion of power at low frequencies increases with the persistence of serial correlations. A 'white noise' series (such as an i.i.d Gaussian noise series) lacks dependence between observations and is characterised in a time series plot as high-frequency jagged spikes. White noise contains equal power at all frequencies, resulting in a linear regression with zero slope ($\alpha = 0$). A 'brown noise' series contains highly correlated observations, with the correlations decaying slowly with increasing lags, characterized in a time series plot as slow rolling waves. The long-range correlations result in spectral power dominating at low frequencies, producing a strong negative slope ($\alpha = 2$). A 'black noise' series exists at the extreme of persistent correlations, with very highly correlated observations that persist for extremely long periods. The flood levels of the Nile River, for example, demonstrate black noise ($\alpha \approx 2.8$), with low or high flood levels persisting for centuries before changing (Ward, 2002). Pink, or 1/*f*, noise exists midway between white noise and brown noise. 1/*f* noise contains both short- and long-range serial correlations, characterized by a spectrum that has power concentrated at both high and low frequencies; however, since the long-range correlations predominate, power is greater at lower frequencies than at high. For pure 1/*f* or pink noise, $\alpha = 1$; however, in practice, a more liberal definition of 1/f noise allows for α in the range 0.5 to 1.5.

Long-range dependence is not unique to 1/f noise. A number of other correlated noise processes, such as brown noise, share the property of long-range dependence. What makes 1/*f* noise particularly attractive is that it possesses a number of additional, interesting properties. For example, 1/*f* noise is scale-invariant. Regardless of the time-scale used to view it, a 1/*f* noise process will possess the same statistical properties. For example, a 1/*f* noise process will have the same statistical properties whether the observation interval is seconds, minutes or hours. Alternately, from a probabilistic perspective, 1/*f* noise exists midway between determinism and randomness. White noise represents randomness, since all values are uncorrelated, or independent. Brown and black noises are closer to determinism, since all values are highly correlated, or dependent. 1/*f* (pink) noise falls midway between white and brown, a balance of randomness and determinism.

Interest in 1/*f* noise can also be attributed to its apparent ubiquity in an ever-growing range of fields. It has been identified in the fluctuations of stock prices, populations, Internet traffic, insulin levels, ECG levels, ocean temperature, solar bursts, music, traffic flow, tics in Tourette syndrome, prime numbers and Japanese scripts (Beran & Mazzola, 2001; Campbell & Jones, 1972; Choi & Lee, 1995; Fraedrich & Blender, 2003; Halley & Kunin, 2000; Kulessa, Srokowski, Drozdz, 2003; Liu, 2000; Peterson & Leckman, 1998; Ryabov, Stepanov, Usik, Vavriv, Vinogradov, Yurovsky, 1997; Saiki, Kitagawa & Hayashi, 1999; Veres, Kenesi, Molnar & Vattay, 2000; Wolf, 1997), amongst others.

Given its apparent ubiquity, advocates of 1/*f* noise promote it as a *de facto* universal feature of natural systems. As a result, since the 1990s, “looking for 1/*f* noise” in experimental data has become as popular as “looking for chaos” did in the 1980s (e.g., Gleick, 1986).

1/f Noise and Nonlinear Dynamics

Several researchers (e.g. Clayton & Frey, 1997; Gilden et al., 1995) have claimed that 1/*f* noise is a signature of nonlinear dynamics – more specifically, a signature of “chaos.” Although 1/*f* noise has been shown to arise in chaotic systems (e.g. Handel & Chung, 1993; Shuster, 1996), 1/*f* noise is not a unique property of chaotic systems, particularly when 1/*f* noise is identified in a finite series. For example, Pressing (1999) reviewed two simple stochastic and linear methods of producing 1/*f* noise: multi-scale randomness, and a second-order autoregressive model. Pressing showed that the sum of multiple i.i.d white noise processes operating at (at least) three time-scales can produce a series whose power spectrum resembles 1/*f* noise. Pressing also showed that the second-order linear autoregressive model, $A_{n+1} = (1-\alpha)A_n - \beta A_{n-1} + (C_n - P) + (M_{n+1} - M_n)$, where α and β are first and second order error correction parameters, C_n and M_n are white noise sources, and P = mean(C), produces a series with a power spectrum with a negative slope over several orders of magnitude at low frequencies when α = -β. Although 1/*f* noise and nonlinear dynamics can be associated phenomena, evidence of 1/f noise does not guarantee the existence of nonlinearity, nor does nonlinearity guarantee the existence of 1/*f* noise. It is necessary to identify each individually.

1/f noise in Cognition

The discovery of 1/*f* noise in cognition is commonly attributed to Gilden, Thornton and Mallon (1995), who examined serial correlations in temporal production, spatial production, and simple reaction time tasks using standard power spectral density analysis (PSD). In the temporal production task, participants were provided with a one-minute sample of a target interval (300ms, 500ms, 1s, 1.5s, 5s, or 10s) produced using a metronome, then attempted to reproduce the target interval on 1000 trials[1] by pressing the spacebar on a computer keyboard when they believed the interval had elapsed. Each response also signified the beginning of the interval for the next trial. In the spatial production task, participants aligned the tip of a digital pen with a cross-hair at the centre of a digital tablet, and for 1000 trials pressed the tip of the pen down at a point that they estimated to be one inch away from the fixation point. In the simple reaction time (RT) task, a signal was presented at random intervals on a computer screen, and for 1000 trials participants responded as quickly as possible by pressing the space bar.

---

[1] With the exception of the 10s condition, reduced to only 400 trials.

Fluctuations in the temporal and spatial production tasks produced a curvilinear power spectrum, with of a negative slope ($\alpha \approx 1$) at low frequencies (below 0.2Hz in the temporal production task, and below 0.1Hz in the spatial production task), and a flat slope ($\alpha \approx 0$) at high frequencies. Gilden et al. (1995) interpreted these spectra as evidence of a two-component system, consisting of a cognitive component in the form of 1/*f* noise at low frequencies, and a motor component in the form of white noise at high frequencies. Since 1/*f* noise was found in the production tasks but not in the simple RT task, they concluded that 1/*f* noise could be attributed to the magnitude judgement process involved in the production tasks.

In light of these results, Gilden et al. (1995) proposed a model of RT fluctuations based on a modified version of the Wing and Kristofferson (W-K) model (1973) of fluctuations in simple RT. In the original W-K model, timing error on trial *n* was attributed to the sum of a cognitive component C_n and a motor component M_n,

$$error_n = C_n + M_n$$

where C_n and M_n were both characterised as white noise processes. Gilden et al. modified the W-K model by characterising the cognitive component C_n as $1/f^{\alpha}$ noise,

$$error_n = (1/f^{\alpha}) + \beta(1/f^{0})$$

were β is a constant reflecting the proportion of white noise in the series.

Gilden (1997) provided further evidence of 1/*f* noise in response times for a variety of psychological tasks, including mental rotation, lexical decision, serial visual search, and parallel visual search. As with the temporal and spatial production tasks, Gilden reported curvilinear power spectra for each task, again interpreted as a 1/*f* noise component at low frequencies and a white noise component at high frequencies. Gilden, therefore, expanded the characterisation of the cognitive component C_n to encompass correlated noises "associated with processes of perception, discrimination and choice" (p. 299). In an attempt to provide a theoretical grounding for these results, Gilden suggested a nonlinear dynamical framework. He reported an attempt to test for the presence of nonlinearity in his data using surrogate series testing based on three measures of nonlinearity (correlation dimension, false nearest neighbours, and nonlinear predictability). However, he provided no details about the results, only mentioning that "surrogates and data were indistinguishable" (p.299) for the three measures[2], which he interpreted as indicating no nonlinear dynamical structure. Given the intricacies of performing and interpreting nonlinear dynamical analysis, and its importance for Gilden and colleagues' developing theoretical framework, his brief treatment was inadequate. Yet, despite reporting no evidence of nonlinearity in the data, Gilden continued to develop a nonlinear framework to explain 1/*f* noise in these tasks. Specifically, he suggested that self-organised systems operating on the threshold between order and chaos were responsible for the observed 1/*f* noise.

The basis of the two-component model was Gilden et al.'s (1995) characterisation of fluctuations in temporal and spatial production tasks as a combination of 1/*f* and white noise, and fluctuations in simple RT as white noise. In the temporal production task, as the tempo of responses increased, white noise became more dominant and 1/*f* noise became negligible. Also, as the intervals decreased, the temporal production task more closely resembled a simple RT task. That is, the simple RT task

---

[2] Surrogates and surrogate data testing will be defined in detail in a later section.

can be considered a temporal production task with minimal interval length. The simple RT task, therefore, represented baseline perceptual motor processing, characterised as white noise, while the temporal production task represented a baseline motor response with an added cognitive component resulting from controlled processing (Shiffrin & Schneider, 1977), characterised as $1/f$ noise.

The characterisation of performance in the simple RT task as a white noise process has been questioned (e.g., Newell & Slifkin, 1998; Riley & Turvey, 2002). Several studies have reported evidence of $1/f$ noise in the fluctuations of simple RT. Van Orden, Moreno and Holden (2003) examined fluctuations in a simple RT task, in which ten participants made speeded verbal responses (saying “ta”) to signals presented on a computer monitor for 1100 trials. They then performed the same power spectral density analysis (PSD) used by Gilden et al. (1995), and also performed surrogate series testing against white noise surrogates to determine whether the series were white noise. They found that series from the ten participants had a mean spectral slope of -.66, with slopes ranging from -.3 to -1.0, and all series were significantly different from the white noise surrogates. Wagenmakers, Farrell and Ratcliff (2004) also examined fluctuations in 1024 trials of a simple RT task, requiring participants to press the “?” key on a computer keyboard in speeded response to a signal presented on the computer monitor. They performed standard spectral analysis, as well as fractional ARIMA (ARFIMA) analysis, a technique used extensively in econometrics to test for long-range dependence (see Appendix A for details). Spectral analysis of the fluctuations yielded an average linear regression slope of -.3, while analysis using fractional ARIMA (ARFIMA) analysis yielded evidence of 1/f noise in 17 of 36 series. These and other reports of long-range dependence in simple RT stand in contrast to Gilden et al.'s characterisation of simple RT as producing white noise fluctuations.

Likewise, Van Orden et al. (2003) criticised Gilden et al.'s (1995) characterisation of the 1/*f* noise component as being the result of controlled processing. They examined fluctuations in a speeded word naming task, since familiar word naming, like simple RT, is regarded as an automatic processing task (Shiffrin & Schneider, 1977). For 1100 trials, twenty participants were presented with four and five letter words randomly selected without replacement from a set of 1857 words, with participants pronouncing the word as fast as possible after presentation of the word on a computer monitor. Spectral analysis of the fluctuations in response times revealed linear regression slopes ranging from -.14 to -.49, while surrogate series testing revealed all series were significantly different from white noise surrogates. Although not in the 1/*f* noise range, the series cannot be characterised as white noise.

These findings raise questions about the validity of Gilden et al.'s (1995) two-component model of sequential dependencies in RT. Gilden et al. characterised automatic cognitive processes as white noise, and controlled cognitive processes as 1/*f* noise. Yet, fluctuations in motor control are themselves commonly held to consist of additive deterministic motor M_n and stochastic components (Riley & Turvey, 2002), and has been attributed to various processes, including low-dimensional chaos (e.g. Mitra, Riley, & Turvey, 1997) and correlated noise (e.g. Chow & Collins, 1995). Empirical evidence is, therefore, inconsistent with Gilden and colleagues' characterisation of the cognitive and motor components, since automatic cognitive processes can also produce 1/*f* noise, and controlled cognitive processes can also produce white noise.

Although not widely acknowledged in the recent literature, earlier Japanese studies (Musha, Katsurai, & Teramachi, 1985; Yamada, 1993, 1996) of temporal interval production in tasks closely resembling those of Gilden et al. (1995) also provided evidence of 1/*f* noise. Musha et al. compared synchronised tapping with continuation tapping (which they termed "free tapping") for inter-tap intervals between 300ms and 500ms. Participants held a castanet in the palm of one hand and tapped it with their other hand. Taps were detected using a micro-switch in the castanet, with inter-tap intervals (ITI) recorded to magnetic tape. In the synchronised tapping condition, participants attempted to synchronise the tapping of the castanet with the ticking of a metronome. In the continuation condition, participants listened to the ticking of a metronome for 10 seconds, after which the metronome was stopped and the participants attempted to produce 1600-1700 taps at the same tempo. Participants used the same tempo in both conditions. The tempos used by Musha et al. (300-500ms) are equivalent to the two shortest temporal intervals used by Gilden et al.: 300ms and 500ms. Furthermore, an analogue to their synchronised tapping condition is provided by Gilden (1997), who reported in a footnote that RT latencies for a synchronisation task yielded a power spectrum with zero slope.

In the synchronisation condition, Musha et al. (1985) reported two distinct spectral patterns among participants. For two of the five participants, the spectral slope was zero at all frequencies, consistent with Gilden (1997). However, for the other three participants, the spectrum for each subject had a slope of zero below 0.01Hz, a positive slope between 0.01Hz and 0.1Hz, and a zero slope above 0.1Hz. In the continuation task, results were almost identical to those reported by Gilden et al. (1995). Power spectra had a slope of zero above 0.1Hz, and a slope of approximately -1 below 0.1Hz. Anticipating Gilden et al. by a decade, Musha et al. interpreted these results as evidence of a 1/*f* noise component at low frequencies, and a white noise component at high frequencies.

The only discrepancy between the power spectra of Musha et al. (1985) and Gilden et al. (1995) for the continuation tapping task is the critical frequency, the frequency at which the '1/*f* noise' component ends and 'white noise' component begins. Gilden et al. reported that the 'white noise' component occurs above 0.2Hz, whereas Musha et al. reported that the 'white noise' component occurs above 0.1Hz. Gilden et al. does, however, report a critical frequency of 0.1Hz for the production of spatial intervals. This may indicate that the power spectra for continuation tapping show large individual differences, or that the castanet-tapping task used by Musha et al. is more analogous to the spatial production task used by Gilden et al.

Musha et al. (1985) tested a further ten participants, yielding the same results - with one exception. The power spectrum of one participant had a slope of approximately 2 below 0.1Hz, which they interpreted as a 'brown noise' component rather than a '1/*f* noise' component. This participant was 50 years old and had many years of Zen training (the other participants did not). To determine whether the difference in results could be attributed to Zen training, Musha et al. tested two Zen Buddhist monks with four years of Zen training. One of the monks had a slope of 1 below 0.1Hz, while the other had a slope of 2 below 0.1Hz. However, these results were confounded by the fact that one of the monks also chanted Buddhist scriptures while performing the task. Since a larger, controlled study has not been performed, the relationship between Zen training and performance on continuation tapping tasks remains intriguing but unclear.

There does, however, appear to be a relationship between performance on the synchronisation task and musical ability. The three subjects in Musha et al. (1985) whose data yielded distinctive power spectra in the synchronisation task were all experienced musicians, whereas those whose data yielded white noise had no musical experience. While there is no apparent effect of musical training on performance in the continuation task in Musha et al. (1985), an effect was seen in a follow-up study by Yamada (1996), discussed in detail later.

A final contribution of the Musha et al. (1985) study is anecdotal feedback provided by the participants concerning their subjective experience of the task (see Appendix B for details). If the Musha et al. study is a valid analogue to the Gilden et al. (1995) study, the experience of participants in the Musha et al. study could provide insight into the experience of participants in the Gilden et al. study. For example, a typical comment, from participant 1: "I was relaxed during [synchronised tapping] because I had the ticking standard to follow, but was very tired after the [free tapping]." It is possible that the power spectra for the continuation task may reflect the effects of fatigue rather than intrinsic fluctuations in temporal judgement. The lack of such an effect in the synchronisation task can be attributed to the assistance of an external reference signal (metronome). It is possible that Zen training affected performance by helping participants resist the effects of fatigue.

In a follow-up to Musha et al. (1985), Yamada (1996) examined inter-tap intervals at three fixed tempos (180, 370, and 800ms) and at an arbitrary (but consistent) tempo, where participants were asked to tap at any tempo they felt comfortable maintaining for the duration of the session. What distinguishes Yamada's study is the test duration. Participants undertook four experimental sessions, one for each tempo, with each session lasting three hours. Sessions were conducted contiguously, with a 3 minute break between blocks of 720 trials, and a 20 minute break between sessions of 5 blocks, resulting in contiguous test duration of 12 hours. In comparison, test duration for Musha et al. (1985) and Gilden et al. (1995) was approximately 3 hours[3]. If fatigue affects performance, as reflected by the power spectrum, then these effects should be most apparent in the Yamada study.

Spectral analysis of inter-tap intervals revealed three striking results. First, the low frequency region was consistent with Musha et al. (1985) and Gilden et al. (1995), characterised by a linear fit with a slope of approximately -1. For the faster tempos (180, 370ms) the high frequency region had a distinct positive slope, and in one participant a parabolic trend (see Figure 10). These results are inconsistent with Gilden et al.'s characterization of the high frequency region as white-noise.

Second, as is evident in Figure 10, there is a disparity in critical frequencies between tempo conditions. Initially, Yamada suggested that the critical frequency was related to the tempo of the tapping. However, when Yamada plotted the power spectra as a function of frequency in cycles, rather than as a function of frequency in Hz (Figure

---

[3] Details of test duration are provided by Musha et al. (1985) , but not by Gilden et al. (1995). Test duration for Gilden et al. is an estimate, calculated from the task requirements, by the current author.

11), it was apparent that the breakpoint was at 10 cycles, or 20 taps, regardless of the tempo.

Third, individual differences were evident in the obtained power spectra. For the high frequency region, some participants yielded a flat slope, while others yielded a positive slope (Figures 10). For the low frequency region, participants with experience in tapping tasks, or in one case experience as a jazz piano player, had shallower slopes than other participants. Yamada (1996) concluded that individual differences in motor control - specifically, expertise in motor control - affected the dynamics of rhythmic tapping.

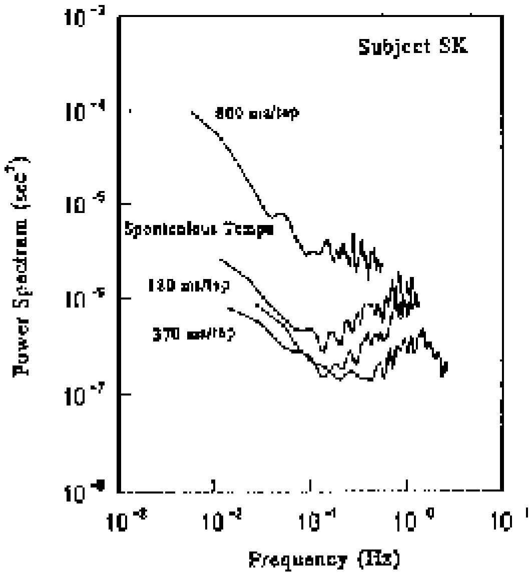


Figure 10. ***Power spectrum of rhythmic tapping at tempos of 180ms, 370ms, 800ms and 'preferred' tempo (from Yamada, 1996)***

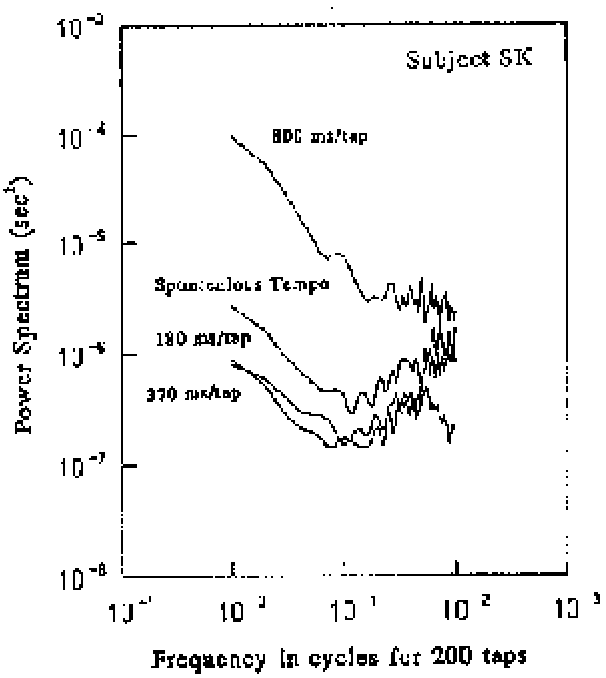


Figure 11. Power spectrum for the four rhythmic tapping tempos, with frequency in cycles (from Yamada, 1996)

An earlier study by Yamada, Imura and Takabatake (1993) found similar differences in power spectra between beginner and experienced musicians performing a rhythmic tapping task at a tempo of 250ms. Spectral slopes for all participants had negative values below the critical frequency, and positive slopes above the critical frequency. However, the beginning musicians had significantly steeper positive and negative slopes than the more experienced musicians. Yamada et al. also suggested that the critical frequency may have been different between the two groups. However, this effect was not clear enough to be definitive. Yamada et al. interpreted these results as evidence that differences in magnitude of variance at all frequencies was related to differences in muscle control, with variance decreasing with increasing experience. However, these results may also indicate differences in ability to sustain attention, as more experienced musicians would posses more experience at sustaining attention.

The effect of musical experience was also addressed by Pressing and Jolley-Rogers (1997) in a study of rhythmic tapping, which compared performance between a musician with extensive musical experience and a non-musician, using power spectral density analysis. Participants performed synchronized tapping over a range of five tempos: 250, 375, 500, 750 and 1000ms. The participant with extensive musical training also performed the task at much shorter tempos: 100, 125, 150 and 175ms. Participants performed 950 trials at tempos below 350ms, and 600 trials at tempos above 350ms. Spectral analysis yielded flat power spectra at lower frequencies, consistent with an autoregressive process rather than a 1/f noise process. Specifically, the best fitting model was a second-order autoregressive model for tempos shorter than 150ms, and a first-order autoregressive model for tempos greater than 175ms. Linear regression of power spectra at higher frequencies also showed that, for each tempo, the slopes of power spectra were steeper for the non-musician in comparison to the musician, consistent with the findings of Yamada (1993, 1996). Furthermore, the power spectra for the experienced musician performing the high frequency tempos (100, 125, 150, 175ms) also showed positive slopes above the critical frequency, similar to those in Figure 10 and 11, inconsistent with the white noise characterization of the high frequency region.

Pressing et al. (1997) speculated that Gilden et al.'s (1995) identification of 1/*f* noise rather than an autoregressive process was due to their failure to remove a quadratic trend in the data prior to analysis, resulting in spurious 1/*f* noise, due to the effects of medium- and long-range trends. They therefore attempted to replicate Gilden et al.'s task by having the participant with musical experience perform 2000 trials of a synchronisation tapping task at tempos of 250ms and 750ms (also consistent with two of the tempos used by Musha et al., 1985). As anticipated, the resulting time-series were non-stationary. Pressing et al. compared the results of spectral analysis for the original, non-stationary series, and for the series following quadratic detrending. In the raw series, they reported an increase in power at low frequencies, with a steeper slope in the 750ms task, consistent with Gilden et al's findings. However, after quadratic detrending, there was substantially less power at low frequencies. Pressing et al. therefore argued that the 1/f noise reported by Gilden (and, by extension, Musha et al., 1985; Yamada, 1996; Yamada et al., 1993) was spurious, resulting from non-stationarity of the data series.

Wagenmakers, Farrell and Ratcliff (2004) used ARFIMA modelling (described in detail later) to simultaneously test for evidence of short-range autoregressive (ARMA) processes, as found by Pressing et al. (1997), and 1/*f* noise, which can be characterized as a type of long-range dependence. They examined data from six participants performing a simple RT, choice RT, and temporal estimation tasks in a single session. In each task, a digit between 1 and 9 was presented on a computer monitor. In the simple RT task, participants were required to immediately press a response key after the appearance of the stimulus. In the choice RT task, participants were required to press one of two keys mapped to the odd and even digits, respectively. In the estimation task, participants were required to press a response key 1s after the appearance of the stimulus. For each task, short- and long-RSI (response to stimulus interval) conditions were also used. Wagenmakers et al. found evidence of both short and long-range dependence. Evidence for long-range dependence was weakest in the simple RT and choice RT tasks. However, stronger evidence for long-range dependence in general, and 1/*f* noise in particular, was found in the estimation task, specifically in the long-RSI condition of the estimation task.

In summary, evidence of 1/*f* noise has been reported in data from a variety of experimental paradigms (e.g., Gilden et al., 1995; Musha et al., 1985, Yamada, 1993, 1996; Wagenmakers et al., 2004). However, this evidence has been criticised on both theoretical and methodological grounds. From a theoretical perspective, Van Orden et al. (2003), for example, disputed Gilden et al.'s characterisation of 1/*f* as being associated with controlled processes. They demonstrated that 1/*f* noise can occur in a task involving automatic processing, and is therefore not unique to controlled processes. From a methodological perspective, Pressing et al. (1997) examined fluctuations in a rhythmic tapping task similar to Gilden et al.'s interval estimation task and showed that the power spectrum, although superficially similar to that for 1/*f* noise, is better fit by an AR(1) or AR(2) model. It is possible that Pressing et al.'s failure to identify 1/*f* noise in their data was due to the brevity of their task. For sessions of 950 trials with intervals between 350ms and 1000ms, each session would last approximately 4 to 15 minutes, respectively. These sessions are sufficiently brief that fatigue should not affect performance. Further, Pressing et al. used a synchronization tapping task, which as discussed earlier may be less fatiguing than Gilden et al.'s (1997) continuation tapping task. If 1/*f* noise is associated with fatigue, it would not be surprising that Pressing et al. did not find evidence of 1/*f* noise. Wagenmakers et al. used a shorter session of approximately half an hour, so fatigue effects are less likely, and found evidence for 1/*f* noise, particularly in their estimation task. However, once again this was a continuation task and so may have been quite demanding, so the effects of fatigue cannot be ruled out.

Nonlinear Dynamics in Cognition

In response to the initial evidence of 1/*f* noise reported by Gilden et al. (1995), and Gilden's suggestion that 1/*f* noise could be attributed to nonlinear dynamics, Clayton and Frey (1997) examined fluctuations in choice RT using spectral analysis, and compared the results to the predictions of a simple nonlinear dynamical system: the logistic equation. Twelve participants performed three two-choice RT tasks with differing memory loads. In all conditions, a signal ('X' or 'O') was presented on a monitor, and participants responded by pressing one of two keys. In the first task, a choice RT task (X-O in Figure 12), participants responded to the stimulus presented on the current trial by pressing the key mapped to the stimulus. In the same-different task (SD1 in Figure 12), participants reported whether the signal was the same as, or different from, the one on the last trial. In the third condition, participants reported whether the signal was the same as, or different from, the signal two trials before (SD2 in Figure 12). They reported success in replicating the results of Gilden: 1/*f* noise at low frequencies, white noise at high frequencies. However, the spectral slopes for the reported "1/*f* noise" at low frequencies had slopes of -.46, -.28, and -.26 (Figure 12). Although they are different from zero, none of these slopes are in the 1/*f* range.

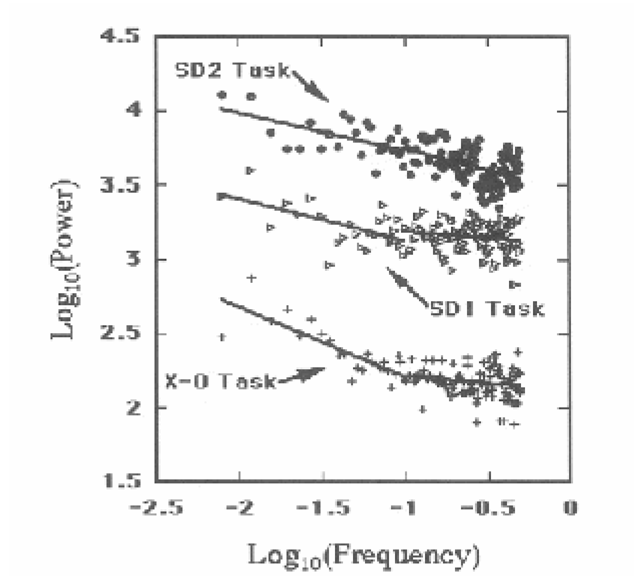


Figure 12. Power spectra for three 2-choice RT tasks: simple 2-choice RT (X-0), same-different lag 1 (SD1), and same-different lag 2 (SD2) (Clayton & Frey, 1997).

Furthermore, the decomposition of spectra into high and low frequency regions above and below 0.1Hz appear to have been determined *a priori* on the basis of Gilden's results. The linear fits of the same-different (SD1) task and 2-choice (X-O) task spectra above and below 0.1Hz result from the *a priori* assumption that such fitting is necessary, whereas it is only necessary in order to replicate Gilden et al.'s results. Although the results of the study indicate the presence of correlated noise in their data, it shows no evidence of 1/*f* noise, nor compelling evidence to support the two component model proposed by Gilden et al. (1995).

Where Gilden et al. (1995) performed ambiguous inferential testing for nonlinear dynamics in their data, which they reported as negative, the extent of Clayton and Frey's nonlinear dynamical analysis consisted of a comparison of power spectra for their own data with those of the nonlinear logistic equation for various values of the control parameter. Specifically, following Gilden et al.'s suggestion that 1/*f* noise reflects a nonlinear system on the edge of chaos, Clayton and Frey examined spectral slopes of a nonlinear system, the logistic map, with parameter values near *r*=3.828327, which produces a chaotic attractor (Schroeder, 1991). Figure 13 shows the time series and power spectra, respectively.

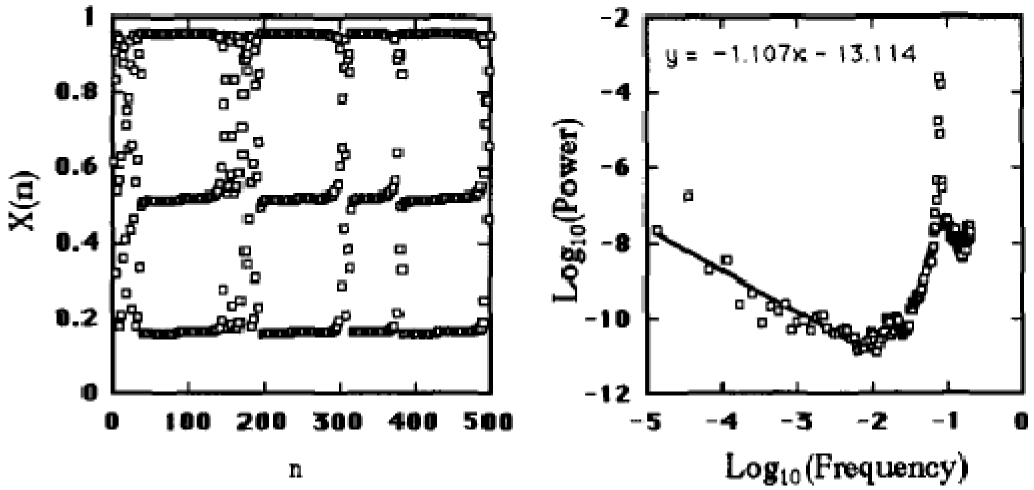


Figure 13. ***Segment of a time series showing the logistic map on the verge of a three-point attractor (left), and the power spectrum of the series showing a negative slope in the low frequency region, and a steep curvilinear slope in the high frequency region (from Clayton & Frey, 1997).***

Linear regression of the low frequency region of the power spectrum yielded a slope of approximately -1, which Clayton et al. reported as pink noise, and interpreted as evidence that a nonlinear dynamical system (the logistic equation) on the edge of a chaotic attractor produces 1/*f* noise. With a control parameter of *r*=4 (Figure 14), the logistic map produces a low-dimensional chaotic series, and for values slightly below this value, for example *r*=3.97 (Figure 14), the logistic map produces a series 'on the edge of chaos,' with the associated power spectrum yielding a linear regression with slope of approximately -.3 over lower frequencies. Since this value is similar to that of the regression slopes of power spectra for their own data, Clayton and Frey concluded that their data show evidence of a nonlinear dynamical system at the threshold of chaos. Although the results of their study show evidence of correlated noise, they do not evidence of 1/*f* noise. Nor does it provide robust empirical evidence of nonlinear dynamical structure in their data.

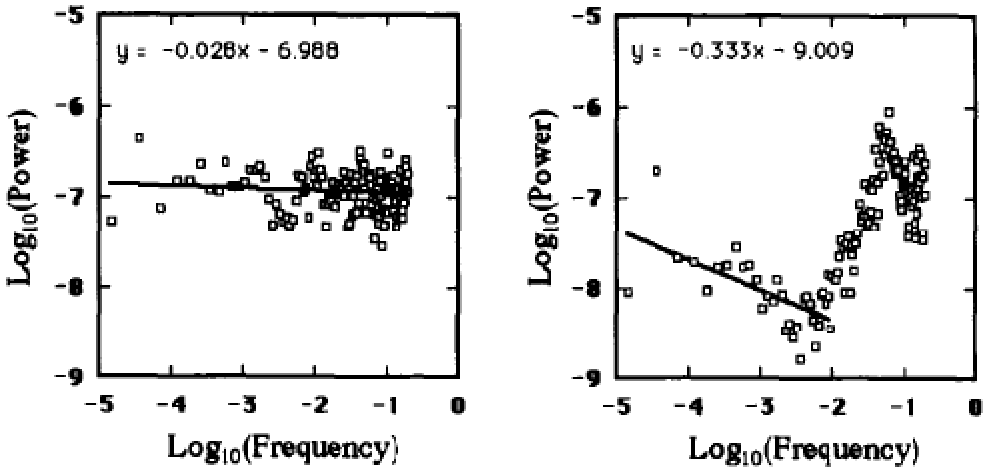


Figure 14. power spectra of the logistic map for r=4 (left) and r=3.97 (right; Clayton & Frey, 1997).

Gilden et al. (1995) reported finding 1/*f* noise in their choice RT tasks, and white noise in choice RT tasks involving task-switching. They interpreted these results in terms of consistent and inconsistent mental set, which produce 1/*f* noise and white noise, respectively. Clayton et al. (1997) interpreted a spectral slope of $\alpha \approx 1$ as indicative of a system at the edge of a three-point nonlinear attractor, and a spectral slope of $\alpha \approx 0$ as indicative of a chaotic system. However, Ward and Richards (2001) argued that the results of Clayton et al. (1997) were confounded by differences between tasks. In response, they replicated Clayton et al.'s study using a set of tasks that differed only in their cognitive and response demands, that is, choice RT tasks with a different number of choices. Specifically, participants performed four tasks, presented in a random sequence. In the first task, participants immediately pressed a key whenever one of two signals appeared; that is, a zero-choice or simple RT. In the second task, participants responded to one of two signals by pressing the key mapped to that signal. In the third task, participants responded to one of four signals by pressing the key mapped to the signal. Power spectral density analysis revealed regression slopes that increased in magnitude with an increase in the number of choices: α=.24 for the simple RT, α=.36 for the two-choice, and α=.61 for the four-choice. Only the four-choice task had slopes in the 1/*f* noise range, and these were at the lower liberal bounds of 1/*f* noise.

A more robust analysis of both 1/*f* noise and nonlinear dynamics was performed by Kelly (2001), and in part reported in Kelly, Heathcote, Heath and Longstaff (2001). Kelly et al. examined choice RT using spectral analysis and nonlinear dynamical analysis techniques. In the first of two experiments, they examined a compatible mapping four-choice RT task. In the first experiment they used a constant inter-stimulus interval (ISI) of one second and in the second experiment ISI was manipulated, with self-paced, forced-pace slow, and forced-pace fast conditions. Participants in all conditions first performed a self-paced session. In the forced-pace conditions subsequent sessions used a fixed ISI, set at the average of the participant's self-paced RT in the fast-paced condition and two standard deviations above that value in the slow-paced condition. Spectral analysis yielded mixed results. Half of the series in the first experiment produced slopes below α=0.5, with the average slope α=0.44, and a range of α=.097 to α=.76. This provided weak evidence of 1/*f* noise at best. These results suggested that there was some correlated noise in the data, but not likely 1/*f*. However, 80% of series had evidence of nonlinear dynamics. In the second experiment, all spectral slopes were significantly less than α=.5. If 1/*f* noise was glimpsed in experiment 1, it was demonstrably absent from experiment 2. Again, these results suggested that there was some correlated noise in the data, but clearly not 1/*f* noise. However, all series in the fixed-pace fast condition, 25% of series in the fixed-pace slow condition, and none of the series in the self-paced condition showed evidence of nonlinear dynamics. The study provides weak evidence of 1/*f* noise in choice RT, but strong evidence of nonlinear dynamical structure related to cognitive and response demands. Experiment 2 also demonstrated that 1/*f* noise and nonlinear dynamics are not necessarily equivalent phenomena. However, it is also possible that the different response mechanisms used by Kelly et al. in experiments 1 and 2 might explain some of the differences in dynamics between the two experiments (Heathcote, 2006, private communication). In the first experiment, responses were made by pressing the relatively insensitive keys of a computer keyboard, while in the second experiment, responses were made by pressing the very sensitive keys of a PS/2 mouse. It is possible that the physical difference in response mechanisms, rather than the difference in task pace, may have contributed to, or masked, the response dynamics in the two experiments.

One explanation offered by Kelly et al. (2001) for the apparent nonlinear dynamics in their fast paced choice RT data is that participants employed a nonlinear dynamical response strategy in order to constrain their response times. Response time tracking was identified by Rabbitt (1968) in choice RT tasks. He found that participants systematically reduced their response times until they made an error, after which they increased their response times for approximately three trials (cf. Laming, 1968), before reducing their response times again. Hence, participants made successive approaches to lower response times that were more likely to result in errors (Rabbit & Vyas, 1970). As a result of this process, participants identified a lower response boundary and adjusted their response times to avoid fast responses near the lower boundary.

Kelly et al. (2001) suggested that participants in their choice RT study employed a similar response time tracking strategy in the forced-pace conditions, and that this strategy may be responsible for the nonlinear dynamics found in the forced-pace conditions. As in Rabbitt (1968), if participants responded too fast, they risked making errors; for example, by anticipating the signal rather than responding to it. However, because the pace was fixed, if participants in Kelly et al. responded too slowly, they also risked not responding before the end of the trial, and in the fast pace condition the upper response boundary was close to the lower response boundary. Kelly (2001) speculated that nonlinear dynamics may have emerged through participants' efforts to produce responses that fell within the narrow band between these two boundaries; that is, that the fixed fast pace condition was effectively an interval estimation task.

Kelly (2001) reported that many of the reconstructed attractors for participants in the fast fixed pace condition resembled the chaotic Lorenz attractor (Figure 4), with a stable fixed-point attractor and two adjacent unstable cycles. Furthermore, unpublished experiments reported in Kelly showed that with practice the reconstructed attractors began to resemble single fixed-point attractors. The two unstable cycles, therefore, were attributed to the two extremes of the response boundary, fast and slow response times, which dissolve with practice as the participant contracts their responses to the middle "safe" response region. Thus, with practice the system undergoes a transition from an unstable chaotic attractor to a globally stable fixed-point attractor. If their results reflect a Lorenz-like nonlinear system with performance demands as a control parameter, then the apparent lack of nonlinear structure in some series would indicate that the task did not represent sufficiently high performance demands for some participants.

In a follow-up to Kelly et al. (2001), Elliott (2003) investigated entrainment of response times as a possible source of nonlinear dynamics and 1/*f* noise in Kelly et al.'s Choice RT task. Kelly (2000) suggested that the Lorenz-like attractors reconstructed from their data reflected entrainment of response times between response boundaries, which prevented participants from responding too fast or too slow. Elliott (2003) therefore investigated RT entrainment in a forced-pace interval estimation task with explicit upper and lower response boundaries.

Interval estimation is an implicit feature of response time tracking and control. What makes it most attractive as an explanation for the dynamical structure in Kelly et al. (2001) is that evidence of both nonlinear dynamics (e.g. Engbert et al., 2002; Ward, 2002) and 1/*f* noise (e.g. Gilden et al., 1995; Wagenmakers et al, 2004) have been reported in interval estimation tasks.

For example, Ward (2002) examined time series produced by six participants in four interval estimation conditions: 1s, 3s, the participant's own "natural" rhythm, and "random" intervals. In each condition, participants were asked to press a key on the keyboard after the respective interval had elapsed. Ward examined the time series produced by each participant for both linear and nonlinear structure, using simple ARIMA modelling and a nonlinear prediction algorithm, respectively. The results of the ARIMA analysis showed that the best fitting model was an ARIMA(0,1,1), or integrated moving-average (MA) model. Ward also used a nonlinear prediction algorithm (explained in detail in a later section) which provided differential diagnosis of i.i.d, linear and nonlinear dynamical structure. The results of the prediction algorithm showed that the series were neither i.i.d or linear, but indicative of noisy nonlinear structure.

The interval estimation task presented by Elliott (2003) involved a simple computer task that required participants to respond at a fixed interval after the appearance of a stimulus on the computer display. To engage participants, the task was presented in the form of a simple video game. The computer display showed a small stick figure, representing the participant, on the left side of the display. The stimulus was a small triangle that appeared on the right side of the display and moved towards the stick figure (Figure 17). On each trial, the participant was given one opportunity to press the response key to make the stick figure jump and land precisely on the centre of the approaching triangular stimulus. In order to land precisely on the centre of the stimulus, the participant had to respond at a fixed interval after the appearance of the stimulus. Narrow upper and lower response boundaries were imposed, with the response key only operating within the allowed response boundaries, thus making explicit the hypothesised boundary conditions suggested by Kelly et al. (2001).

To investigate whether any effects were due to the interval estimation process or due to the pace of the task, the task involved two paces, slow and fast, determined by the inter-stimulus interval (2250ms for the slow pace, and 1500ms for the fast pace) and the response interval. For each pace, there were two further conditions: fixed and random-ISI. In the fixed conditions, the ISI was constant in each condition. In the random conditions, the ISI varied randomly around the fixed ISI. Response times across conditions were standardised by removing the target RT of each condition from the observed values. Analysis was, therefore, performed on the latencies between the measured response time and the target response time. Non-stationarity of the data series was identified by testing for invariance of the mean, standard deviation and joint-probabilities across segments of the series, using Schreiber's (1997) cross-prediction method (described in detail in a later section). This method showed that stationarity was limited to the second half of the series; therefore, analysis was limited to the stationary second half of the data series, which yielded approximately 900 trials for analysis.

Standard spectral analysis of the stationary region of the series produced shallow regression slopes of the bi-log power spectra. The steepest regression slope was α=.26, well outside of the 1/*f* noise region. Nonlinear dynamical analysis consisted of surrogate series testing using nonlinear prediction error and time reversibility indices, in the manner of Kelly et al. (2001). Nonlinear prediction error was significantly less than for the surrogates for two of the sixteen participants, both of whom were in the fixed-pace-fast condition. Time reversibility was significant for seven participants: one from the random-ISI-slow condition, two in the random-ISI-fast condition, and four in the fixed-ISI-fast condition.

These results were consistent with the findings of Kelly et al. (2001), insofar as both studies showed evidence of nonlinear dynamics related to task pace. However, Elliott's results were not nearly as compelling. More importantly, they failed to support the hypothesis that nonlinearity in the data could be attributed to temporal estimation of responses, since not all of the conditions showed evidence of nonlinear dynamics. However, since evidence of nonlinear dynamics was predominantly limited to the fast conditions, the results were consistent with the hypothesis that nonlinear dynamics emerge as a function of performance demands. That is, the slow-pace conditions were sufficiently lacking in task demands to elicit nonlinear dynamics. The results also suggested that the fast-paced conditions were also sufficiently lacking in task demands to elicit nonlinear dynamics for most participants.

One purpose of the current study is to test this hypothesis by further increasing the task demands of the temporal estimation task used by Elliott (2003). This will be achieved in two ways. First, the task demands will be increased by further increasing the pace of the task. Second, task demands will be increased by adding a choice element to the task. Specifically, the choice conditions will involve two- and four-choice tasks. This will also bring the task closer to that of Kelly et al. (2001).

Furthermore, Elliott (2003) also failed to find evidence of 1/*f* noise in interval estimation, inconsistent with the findings of Gilden et al. (1995). One explanation for this inconsistency is the careful measures taken by Elliott (2003) to ensure stationarity prior to data analysis. Gilden et al. provided no evidence of having tested for, or corrected, non-stationarity in their data prior to analysis. As mentioned previously, Pressing and Jolley-Rogers (1997) provide a compelling argument that the 1/*f* noise reported by Gilden et al. is spurious, a result of non-stationarity in their data. The current study will also briefly examine the impact of non-stationarity of data on the results of several tests of 1/*f* noise.

Such inconsistency in results also highlights one of the major issues confronting dynamical analysis of psychological data: the reliability of measures of sequential structure. This issue has been of increasing concern in other fields (see, for example, Eke et al, 2000, 2002), and has recently been raised in regards to analysis of psychological data (e.g. Wagenmakers et al., 2004; Delignieres et al, 2005). Another concern of the current study will be the accuracy of various analysis methods, which will be discussed in the next section. Specifically, the current study will include application of four measures of 1/f noise: standard power spectrum analysis, the competitive ARFIMA modelling framework of Wagenmakers et al. (2005), and two methods proposed by Eke et al. (2001): refined power spectrum analysis, and the dichotomous Dispersion-Scaled Window Variance (Disp-SWV) method.

The aim of the current study, therefore, is two-fold. First, to determine whether increasing the performance demands of a temporal estimation task, by increasing the speed of the task, or by including a choice component (with two and four choices), will yield stronger evidence of nonlinear dynamical structure or 1/*f* noise in the resulting time series data. Second, whether results obtained by application of three recent measures of 1/*f* noise will differ from those provided by application of standard spectral analysis methods. It is anticipated, following Kelly et al. (2001) and Elliott (2003), that the increased performance demands required by the tasks in the current study will yield stronger evidence of nonlinear dynamical structure, in terms of providing evidence of nonlinear dynamics in a greater number of series, as more participants should find the current tasks more demanding than those in the previous study.

Analysis Methods

Fractal Techniques

Fractal analysis techniques address a single question: does the experimental series under investigation contain long-range dependence (Karagiannis et al., 2002). The question is straightforward, but obtaining an answer is not. A variety of methods have been developed in the physiology literature (and more recently applied to psychophysics, see Delignieres, Torre & Lemoine, 2004), but they can provide misleading and contradictory results (Eke, Herman, Bassingthwaighte, Raymond, Percival, Cannon, Balla, and Ikrényi , 2000). The most commonly used methods are evaluated here. Details of each method are provided in Appendix A.

Despite the apparent variety of methods, the most commonly used techniques rely on a common strategy: a statistical measure q is calculated from the experimental series for non-overlapping segments of length n, for increasing values of n (Eke, Hermann, Kocsis and Kozak, 2002). On a bi-logarithmic plot of q versus n, a scaling exponent ε is estimated from the slope of a linear regression of log-q versus log-n. The presence of long-range dependencies is then determined by the value of the scaling exponent ε.

Evaluation of Fractal Techniques

Attempts to the answer the question “Is there correlated noise in the experimental data?” typically involve one of the following fractal analysis techniques being applied to experimental data, then rejection or acceptance of correlated noise based on the value of the estimated scaling exponent. More often than not the technique of choice is power spectral density (PSD) analysis. However, the validity of this approach, and indeed any single approach by itself, has been recently questioned. Eke et al. (2000), Eke et al. (2002), and Karagiannis, Faloutsos, and Riedi (2002) reviewed the accuracy and sensitivity of these techniques (with the exception of the spectral classification method) to physiological data, and found that they can often yield inaccurate and conflicting estimates of long-range correlations. Karagiannis et al. (2002) also reviewed several of these techniques, using artificial series containing long-range correlations (to test accuracy) and artificial series which did not contain long-range correlations but did contain structure (for sensitivity; for example, cosine functions with added noise) and concluded that the techniques can easily be fooled (particularly by periodicity) and that no single technique can accurately identify long-range correlations. Rangarajan and Ding (2000) showed that use of power spectral density analysis can lead to false positives. This brings into question the reliability of previous interpretations of long-range correlations in data based on these methods.

The appropriateness of two more recent approaches, the competitive ARFIMA (Wagenmakers et al., 2005), and the spectral classifier (Thornton & Gilden, 2005), remains unclear. Adoption of new techniques before they have been properly evaluated may result in the same uncertainty that has arisen from premature adoption of the earlier spectral analysis techniques (e.g., Eke et al., 2000). Contributing to the uncertainty is the apparent irreconcilable differences of the authors of the two approaches, whose mutual criticism may undermine confidence in either method. Thornton and Gilden (2005), for example, argue that the competitive ARFIMA approach does little more than "demonstrat[e] that a few experiments in RT are more likely fractal than ARMA" (p.421). Wagenmakers et al. (2005) demonstrated that both approaches are equally effective – although they argue that their ARFIMA method is computationally more efficient – yet remain sceptical about the benefits of fractal analysis in psychology without a supporting theoretical framework (see Van Orden, Moreno & Holden, 2003, for a possible metaphysical framework).

Nevertheless, in response to evaluations of individual fractal analysis techniques (see also Caccia, Percival, Cannon, Raymond, & Bassingthwaighte, 1997; Cannon, Percival, Caccia, Raymond, & Bassingthwaighte, 1997) several strategies involving multiple techniques have been proposed to identify 1/*f* noise in experimental data. Rangarajan and Ding (2000), for example, showed that a single measure, such as PSD, used in isolation can provide misleading results. They therefore advocate the use of a battery of tests, with a view to identifying convergent results.

Another methodological framework, developed by Eke et al. (2000, 2002), involves a multiple-stage process. The first stage is a "pre-analysis" using power spectral density (PSD) analysis, which determines the method to use for the main analysis, based on the results of the PSD analysis. If the spectral analysis method yields a power law relation between spectral power and frequency for a 100-fold range of frequencies, fractal analysis is warranted. For $\alpha=(-1,1)$, the series is best characterised as fractional Gaussian noise (fGn), while for α=(1,3) the series is best characterised as fractional Brownian motion (fBm; Eke et al., 2000). The distinction is crucial, as each class has specific analysis techniques which will yield reliable results for it, but spurious results for the other class. This is due largely to the requirement of stationarity – fGn is a stationary process, whereas fBm is not. However, techniques that require stationarity can be applied to fBm process after the fBm series is differenced, as this will yield a fGn series. For an fGn series, Dispersional analysis (Disp) is the most accurate measure of long-range correlation. For an fBm series, Scaled Windowed Variance analysis (SWV) is the most accurate measure of long-range correlations. Use of either test for the wrong class will, however, yield spurious results. This approach has been adapted to psychophysical data by Delignieres, Torre and Lemoine (2005).

This brief evaluation of fractal analysis techniques highlights the complexity of answering the question: "Does the experimental series contain correlated noise?" Although there are several promising new approaches to answering this question, there is not yet a systematic approach that guarantees a reliable answer. This brings into question the ubiquity of 1/*f* noise in psychological data. As Eke et al. (2000) state: "Most of the published data in the literature has resulted from studies unaware of the issues involved in fractal analysis, and their revision for potentially being in error is warranted." Although their comments are directed towards findings of long-range dependence in physiology, researchers in search of 1/f noise in experimental psychology should also take heed.

Nonlinear Techniques

Evaluation of nonlinear dynamical structure in the data will apply the framework described in Heathcote and Elliott (2005) for the evaluation of nonlinear structure in noisy data, which will be briefly outlined below using results from Elliott (2003) to illustrate the relevant techniques.

Stationarity

Stationarity is crucial for linear and nonlinear dynamical analysis in general, particularly for the fundamental representation used in nonlinear dynamics, a delay embedding. Most nonlinear analysis techniques involve the estimation of measures based on local neighbourhoods (nearby sets of points), defined according to a distance measure in the embedding space (see section after next). A requirement of such measures is stationarity. Linear analysis techniques generally only require 'weak' stationarity, defined as an invariant mean and standard deviation between segments of the series (Kaplan & Glass, 1995). Nonlinear analysis techniques, however, require 'strict' stationarity. Kantz and Schreiber (2003) define stationarity as "all joint probabilities of finding the system at some time in one state and at some later time in another state are independent of time within the observation period, i.e. when calculated from the data." (p. 14). More formally, Casdagli (1997) defined nonstationarity as: "…a time series $x_1, x_2,…,x_N$ is nonstationary if, for low *m*, there are variations in the estimated joint distribution of $x_i, x_{i+1}, …, x_{i+m-1}$ that occur on time scales of order N". (p. 12)

Time series data from biological systems, and psychological systems in particular, are typically nonstationary, so it is usually necessary to eliminate a nonstationary section of the series, such as an initial nonstationary region of the data, or to limit analysis to a smaller stationary segment of the series. One method of determining ‘strict’ stationarity is Schreiber’s (1997) cross-prediction method, which exploits the invariance of joint probability distributions in a stationary series. If a stationary series is divided into sufficiently long segments, invariance of joint probability distributions means that information about one segment of the series should allow accurate prediction of another segment in the series. Poor prediction between segments indicates non-stationarity. Figure 15 shows the application of the cross-prediction method to a series from Elliott (2003). The series has been divided into six equal segments, and the cross-prediction algorithm has been used to predict the structure of each segment using the structure of the other five segments. The first segment poorly predicts the other five segments, but segments 2-6 provide reasonable prediction of each other, indicating initial non-stationarity in the first segment of the series. The series can be rendered stationary by removing the initial non-stationary segment.

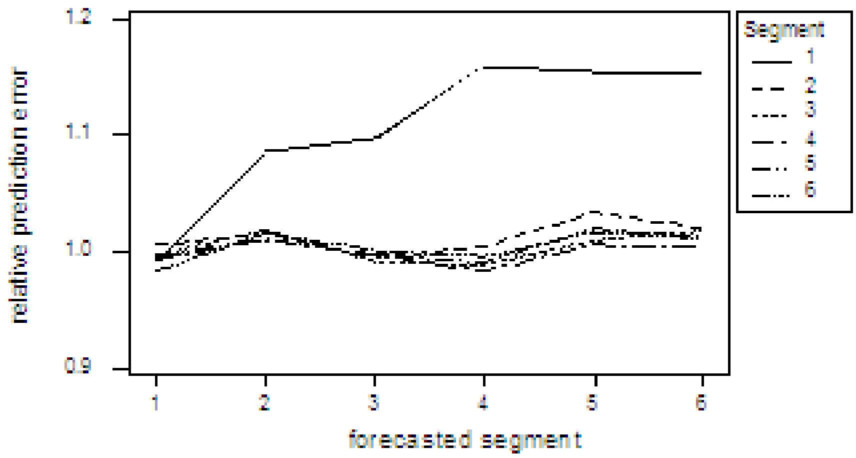


Figure 15. Prediction errors for subject seven in the fast-fixed group. The legend indicates the segment used to derive predictions.

Linear Dynamical Analysis

After resolving the issue of stationarity, linear dynamical analysis provides a useful preliminary to nonlinear dynamical analysis, as nonlinear systems typically produce linear autocorrelations. Linear analysis employs standard linear autoregressive moving-average (ARMA) analysis techniques (Box & Jenkins, 1976), yielding an ARMA(p,q) model that predicts future values from linear combinations of past values and noise. The best fitting autoregressive model is determined by the number of significant partial autocorrelation function values. The partial autocorrelation function also provides a first glimpse of nonlinear structure in data. Nonlinear structure typically results in large partial autoregressive parameter values, consistent with long range dependencies. If a process is nonlinear, then linear models will necessarily fail to explain all of the residual structure in the series.

Embedding

The fractal analysis techniques discussed earlier transform an experimental time series into frequency information. Nonlinear dynamical analysis techniques, on the other hand, transform an experimental time series into geometric information, i.e. a delay embedding in phase space. An *m* dimensional delay embedding transforms a time series into a set of *m* dimensional points ($x_{t+\delta}$, $x_{t+2\delta}$,…$x_{t+m\delta}$), where δ is the time delay between samples. Takens (1981) demonstrated that a one-to-one image of a stationary dynamical system's attractor (the d dimensional set of values it passes through) can be reconstructed from an embedding in a phase space with dimension $m>2d$. The Lorenz system, for example, has an attractor which occupies a subset of three-dimensional space, with $d \approx 2.05$.

Embedding Dimensions and Delays

An optimal delay embedding requires calculation of a time delay δ and embedding dimension De. The time delay (or lag) δ is chosen to optimise the expansion of the attractor in the embedding space. One method for choosing δ is the order of the ARMA(p,q) autoregressive parameter p. Another method for choosing δ is setting it equal the first minimum of the plot of mutual information as a function of time delay (Fraser & Swinney, 1986). The embedding dimension D_e is the Euclidian dimension of space used for the embedding. One method for determining the minimum embedding dimension D_e is the method of false nearest neighbours (Kennel, Brown & Abarbanel, 1992). Heathcote and Elliott (2005) found that when noise is high the shortest available delay should be used.

Nonlinear Prediction

Heathcote and Elliott (2005) found locally averaged nonlinear prediction (Schreiber & Schmitz, 1997) to provide the best available method of testing for nonlinear dynamics in a noisy time series. Nonlinear prediction, first proposed as a test of nonlinear dynamics by Sugihara and May (1990), exploits the phenomena of sensitive dependence on initial conditions, popularly known as "the butterfly effect." Although nonlinear dynamical systems are deterministic, minor differences in initial values see future states rapidly diverge. Nonlinear systems are therefore predictable in the short term, but unpredictable in the long term. Uncorrelated noise, on the other hand, is unpredictable at all time scales.

Nonlinear prediction methods usually use an approach similar to Lorenz's (1969) method of analogues. Lorenz reasoned that if a stationary system is deterministic, similar values at different positions in a series should evolve into similar future values. In order to predict the next value in the series, the prediction algorithm searches the history of values for a similar value, and sees what value it evolves into next. This should provide an accurate prediction of the next value for similar initial values. Sugihara and May proposed a similar method using the evolution of points in the embedding space. To compensate for noise, the algorithm uses the aggregate evolution of neighbouring points rather than a single point, with the aggregation method ranging from simple averaging to more complex combinations.

Inferential Testing

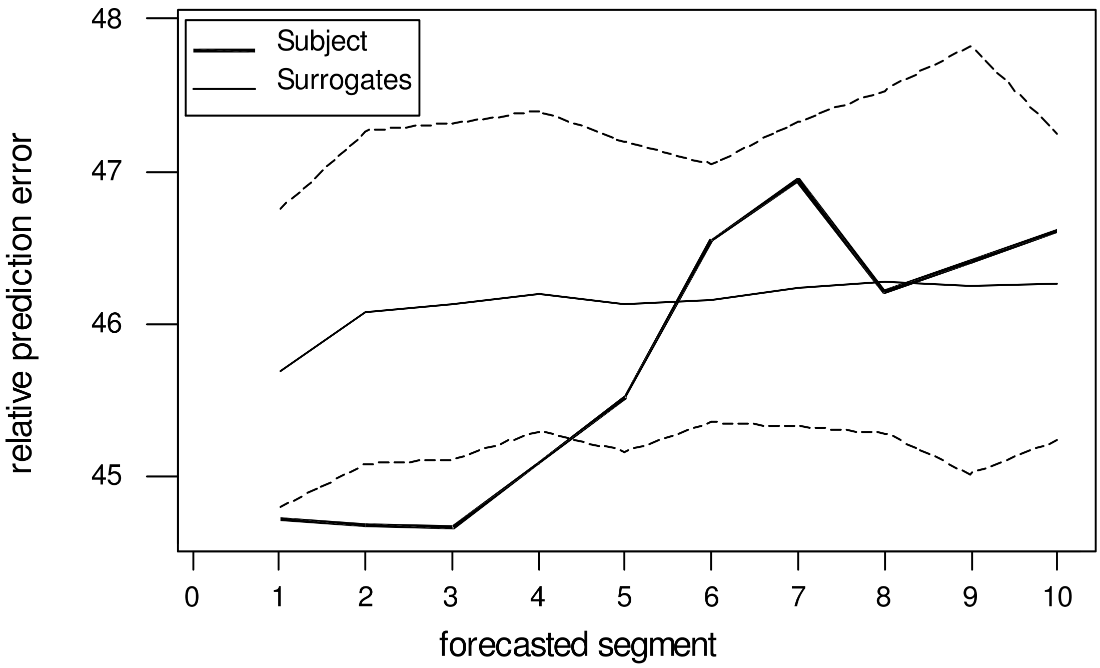
Surrogate series testing (Theiler et al., 1992) provides a method of statistical inference for detecting nonlinearity in time series. Surrogate series are variations of an experimental series that preserve the structure of the experimental series according to a null hypothesis. Random shuffling of the experimental series, for example, produces surrogates that express the null hypothesis of no temporal structure, yet retain the same static distribution of values as the experimental series. In the context of inferential testing for nonlinearity, a common null hypothesis is a linear ARMA process, such as Theiler et al's (1992) Amplitude Adjusted Fourier Transform (AAFT) surrogates, which preserves all short range linear dynamical structure.

Inferential testing involves a comparison of the experimental and surrogate series using an index sensitive to nonlinear structure. The values of measures of the surrogate series act as a distribution of values expected of a series constrained by the null hypothesis, against which the value of the measure for the experimental series is tested. The significance of the test is determined by the percentile of the surrogate distribution into which the experimental measure falls. When nonlinear prediction error is used as the measure, the test is one tailed, with significance corresponding to prediction error that is lower for the experimental than surrogate series.

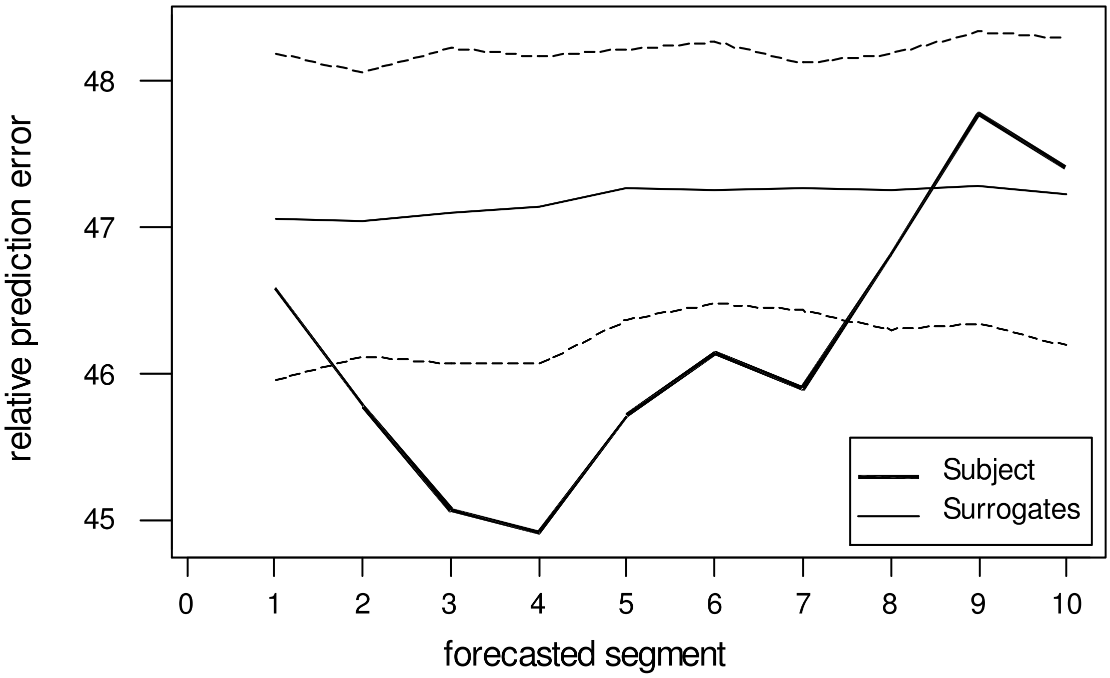
Surrogate series testing of nonlinear prediction errors of series from Elliott (2003) showed that the nonlinear algorithm provided initially lower prediction errors compared to the linear surrogates for series from two subjects in the fast-paced fixed-ISI conditions (Figure 16). For noisy series, Heathcote and Elliott (2005) found this test provided greater discrimination between surrogate series and nonlinear data series for prediction lags greater than one, although the test is almost always applied at lag one (Schreiber & Schmitz, 1997). Figure 16b shows that for lag one, the test does not discriminate between the surrogate series and observed data, but for lags greater than one (and fewer than 8) the test does discriminate, with prediction error substantially lower for the data series than the surrogates.

Present Study

In summary, the present research will address two issues. First, it will examine whether increasing the performance demands of an interval estimation task, by increasing the pace of the task, or by requiring a choice (between two or four options), will produce evidence of nonlinear dynamical structure or 1/*f* noise in the resulting time series data. Second, it will address whether results obtained by application of the most recent fractal analysis methods concur with each other, and with results provided by application of classical and refined spectral analysis methods. We anticipate that the increased performance demands required by the tasks in the current study will yield stronger evidence of nonlinear dynamical structure than found by Elliott (2003), inasmuch as they will provide evidence of nonlinear structure in a greater number of series, since more participants should find the current tasks more demanding than those in Elliott (2003).



(a)



(b)

Figure 16. Relative prediction errors (bold lines) and 95% confidence intervals (dashed lines) for prediction errors of surrogate series for two subjects in the fast-paced fixed-ISI groups (FPF) in Elliott (2003).

Method

Participants

Forty-eight undergraduate psychology students from the University of Newcastle, Australia, provided informed consent to participate in the experiments. Participants were recruited using advertisements placed on a noticeboard in the Psychology building, and advertisements placed on an online undergraduate recruitment noticeboard. Participants received course credit for participating.

Apparatus

The apparatus consisted of an Athlon AMD 1.8GHz personal computer running a Turbo Pascal program using the MS-DOS 7.0 operating system, which presented instructions and stimuli on the monitor and recorded experimental data. Blocks of trials were initiated using the spacebar on the computer keyboard. For the no-choice estimation conditions, responses to stimuli were made using the left button of a two-button mouse. For the two-choice conditions, responses to stimuli were made using the left and right buttons of a two-button mouse. For the four-choice conditions, responses to stimuli were made using the left and right buttons of two modified two-button mice connected to the PC via a dedicated I/O card, whereas responses were collected through a standard PS/2 mouse in the other conditions.

Procedure

Instructions were presented to participants on a single computer monitor screen. The task consisted of a simple computer game (Figure 18), in which "giant bugs" moved horizontally across the screen, slightly above a horizontal line representing a floor, towards the user's ankh-like figure on the left side of the screen. Only one bug appeared on-screen at any time, with the interval between bugs appearing (ISI) varying between the No-Choice condition and the Two-Choice and Four-Choice conditions. Each participant was verbally reminded, after reading the instructions, to respond as accurately as possible, and to avoid missing responses.

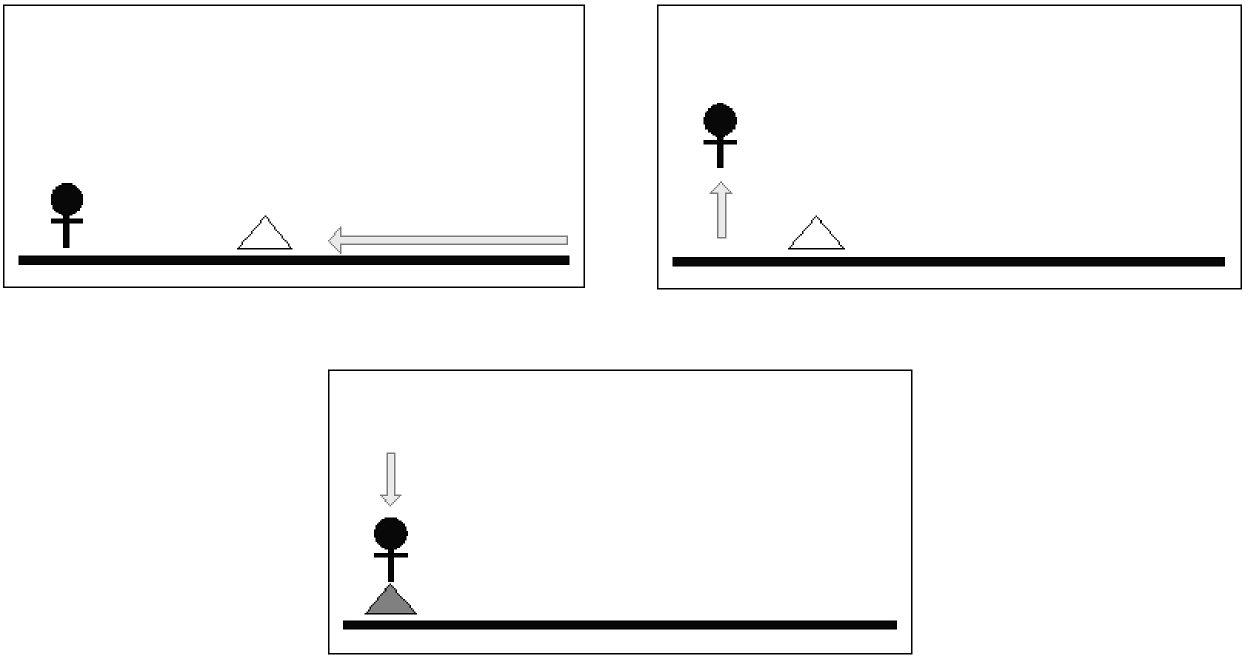


Figure 17. Schematic diagram of the computer task, showing the user graphic, the textured surface beneath, and the "bug" stimuli approaching from the right. Arrows indicate movement of objects. In the upper left panel, the "bug" stimuli approaches from the right.

The participant controlled the ankh on the left side of the display. The participant's response triggered a short animation, showing the ankh 'jumping' vertically at a fixed speed to a fixed height. The ankh landed (i.e. returned to its initial vertical position) after a fixed interval T_{jmp}. On each trial, a triangle representing a 'bug' moved horizontally across the display at a fixed speed towards the ankh. The horizontal centres of the ankh and the triangle aligned at a fixed interval T_{tot} after the appearance of the triangle.

The participant was instructed to make the ankh jump into the air and land precisely on the horizontal centre of the approaching triangle. To achieve this, the participant had to make the ankh jump and return to its initial vertical position when the horizontal centres of the two objects aligned at time T_{tot}. This required the participant to jump at the fixed time T_{jmp} before the two objects aligned. The target response time T_{resp} after the appearance of the stimulus on the screen was, therefore, $T_{resp} = T_{tot} – T_{jmp}$, and the participant's response time RT measures the participant's ability to accurately estimate this interval T_{resp}.

The estimate interval T_{resp} was constant on all trials. In the No-Choice conditions, the estimate interval T_{resp} was 240ms. However, a pilot study of the Two-Choice conditions indicated that participants found it very difficult to perform the Two-Choice conditions at this pace; therefore, the pace for the Two- and Four-choice conditions were reduced and made equal to the fastest pace in Elliott (2003), resulting in an estimate interval T_{resp} of 480ms for the Two-Choice and Four-Choice conditions.

For half of the participants, the ISI was fixed (“Fixed” conditions) at 1500ms, while for the other half of the participants, the ISI varied randomly around the fixed ISI value (“Random” conditions), ranging between 1375ms and 1625ms. The purpose of the random ISI condition was to disrupt the rhythm of the task, thereby preventing participants from developing a rhythmic response, as in a tapping task.

The No-choice, Two-choice and Four-choice conditions differed in the number of possible stimuli and responses. The No-Choice conditions involved a single coloured “bug” stimulus, and required participants to respond by pressing the left mouse button of a two-button mouse. The Two-Choice conditions involved two different coloured “bug” stimuli, with equal numbers of each coloured stimulus being presented in each block in a random order. Each colour was mapped to either the left or right mouse button. The Four-Choice conditions involved four different coloured “bug” stimuli, chosen randomly on each trial with equal numbers in each block. Each colour was mapped to either the left or right mouse button of two modified two-button mice, with participants responding by depressing a different mouse button depending on the colour of the stimulus. The three conditions are summarised in Table 1, which also shows their relation to the conditions in Elliott (2003).

Table 1. ***Summary of the conditions being examined in the current study (No-Choice, Two-Choice, and Four-Choice, in bold), shown in comparison to the conditions previously examined by Elliott (2003).*** **Each cell includes Fixed and Random ISI conditions (not shown).**

	Number of Choices		
Pace	0	2	4
"Slow"	Elliott (2003)		
"Medium"	Elliott (2003)	**Two-Choice**	**Four-Choice**
"Fast"	**No-Choice**		

Response time was recorded as the interval between the appearance of the stimulus on the screen and the depression of a mouse button. However, to standardise the data, analysis was performed on the difference between response time and the estimate interval T_{resp}. For example, if the participant responded 50ms too soon, the transformed score for that trial was -50. For the remainder of this thesis, response times refer to these transformed response times.

A practice block of 50 trials was provided to familiarise participants with the speed of the 'bug' and the duration of the jump. Participants were allowed only one jump per bug. Feedback was provided visually by a colour gradient on the bug (Figure 18), which indicated to the participant how accurately they responded. For a perfect hit (RT=Tresp), 100% of the triangle turned red. As the response became less accurate, a lesser proportion of the triangle turned red. When the participant missed the triangle entirely, the stimulus remained its original colour.

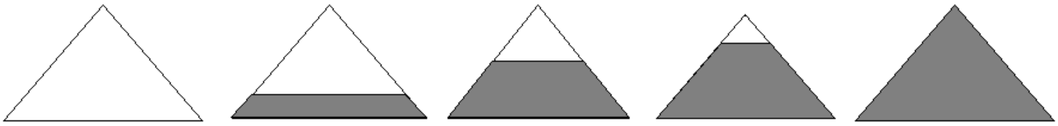


Figure 18. A grey-scale diagram of visual feedback provided to participants, a bi-colour gradient ranging from a miss (left) to a precise hit in the centre of the object (right).

Participants took part in two sessions, lasting approximately 45 minutes each, with each session consisting of 10 equal blocks of 100 trials. This yielded 2000 observations for each participant. Short breaks of around 30s. were provided between blocks, and a 15 minute break was provided between sessions, to reduce fatigue.

Results

Task Demands

For each participant, response times from 20 contiguous blocks of 100 trials were concatenated to produce individual time series: time-ordered vectors of 2000 raw scores. The first 10 blocks (or 1000 scores) were obtained during the first experimental session, while the last 10 blocks (the last 1000 scores) were obtained during the second experimental session. Figure 19 shows a sample time series plot of raw scores for an individual participant.

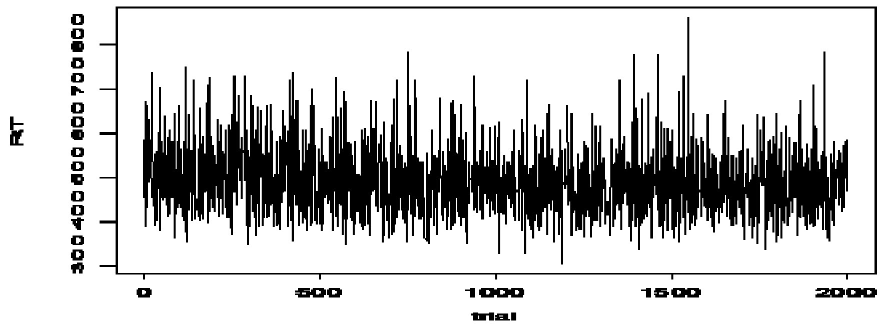


Figure 19. Plot of typical time series from one participant.

To determine whether the three tasks differed in performance demands, four measures were compared across the tasks: the number of missed responses, the number of incorrect responses (stimulus-response mismatch) in the Two- and Four-Choice tasks, the estimation accuracy (discrepancy between average response time and target response time) and precision (standard deviation of response times). These measures were also compared between the two sessions to determine whether they were affected by practice.

The proportion of missed responses in each experiment is summarised in Table 2. The percent of missed responses was compared using a 3 x 2 ANOVA, with Experiment and Session as factors. The decrease in the proportion of misses between sessions was significant, $F(1,45)=37.54$, $p<.01$. There was also a significant difference in the proportion of misses between experiments, $F(2,45)=3.36$, $p<.05$, but no interaction between Experiment and Session, $F(2,45)=.54$, $p=.58$.

Table 2. Percent of missed responses for each session in the three experiments.

	Session 1	**Session 2**
No Choice	1.91	1.19
Two Choice	2.41	1.35
Four Choice	1.77	0.96

A Tukey post hoc test found no difference in the proportion of missed responses between the No- and Two-Choice experiments, $p=.23$, nor between the No- and Four-Choice experiments, $p=.64$, but did find a difference between the Two- and Four-choice experiments, $p=.03$. The proportion of missed responses was highest in the Two-Choice experiment (M=1.88, SD=.88), and lowest in the Four-Choice experiment (M=1.36, M=.71), while the No-Choice experiment (M=1.55, SD=.99) was intermediate, but not significantly greater or lesser than either. In terms of missed responses, the only effect seen is a counter-intuitive inverse relationship between the proportion of errors and the number of choices.

For the Two-Choice and Four-Choice estimation tasks, performance demands were also assessed by comparing the number of incorrect responses (mismatch between the stimulus type and response type). The proportion of incorrect responses in the Two- and Four-Choice estimation tasks are summarised in Table 3, and were compared using a 2 x 2 ANOVA, with Experiment and Session as factors. The increase in the number of incorrect responses between the Two- and Four-Choice experiments was significant, $F(1,30)=37.89$, $p<.01$. However, there was no significant difference in the number of incorrect responses between sessions, $F(1,30)=3.31$, $p=.07$, nor was there any significant interaction between Experiments and Sessions, $F(1,30)=.61$, $p=.44$. This suggests that the number of incorrect responses reflects task-specific demands that are unaffected by practice.

Table 3. Percent of incorrect responses for each session in the Two- and Four-Choice experiments

	Session 1	**Session 2**
Two Choice	9.47	8.2
Four Choice	22.53	22.03

Task demands were also assessed by comparing the means and standard deviations of response times, as measures of accuracy (discrepancy between average response time and target response time) and precision (the amount of variance in response times), respectively. Figure 20a illustrates the mean RT of each session for the three experiments, which were compared using a 3 x 2 ANOVA, with Experiment and Session as factors. Figure 20a shows a significant decrease in mean RT between sessions, $F(1,45)=65.04$, $p<.05$, as well as a significant difference in RT between experiments, $F(2,45)=34.41$, $p<.05$. However, there was no interaction between experiments and sessions, $F(2,45)=.51$, $p=.59$.

As suggested by Figure 20a, a Tukey post hoc test of mean RT across experiments found no significant difference between the No- and Two-Choice experiments, $p=.74$, but did find a significant difference between the No- and Four-Choice experiments, $p<.01$, and between the Two-Choice and Four-Choice experiments, $p<.01$. Participants in the No-Choice and Two Choice experiments performed with remarkable accuracy, on average over-estimating the target interval by 6ms and 10ms, respectively. Participants in the Four-Choice experiment over-estimated the target interval by 54ms, a discrepancy five times greater in magnitude than those in the Two-Choice experiment, yet still demonstrating considerable accuracy.

Figure 20b illustrates the standard deviation of RT of each session for the three experiments, which were also compared using a 3x2 ANOVA, with Experiment and Session as factors. The ANOVA revealed a significant decrease in the standard deviation between the first and second session, $F(1,45)=168.67$, $p<.01$, as well as a significant difference in standard deviations between experiments, $F(2,45)=8.5$, $p<.01$, but no interaction between experiments and sessions, $F(2,45)=3.01$, $p=.06$.

A Tukey post hoc test revealed a significant difference in standard deviations between the No- and Two-Choice experiment, $p=.001$, and between the Two-Choice and Four-Choice experiments, $p=.008$, but found no difference between the No- and

Four-Choice experiments, p=.728. The precision of response times is therefore equal in the No- and Four-Choice experiments, but much lower in the Two-Choice experiment.

In summary, participants in the Two-Choice experiment missed a greater proportion of responses, while participants in the Four-Choice experiment responded incorrectly on a greater proportion of responses. Participants in the Four-Choice task tended to over-estimate, while those in the No- and Two-Choice tasks tended to have near-perfect accuracy; however, participants in the Two-Choice task showed much less variation in response times than those in the No- and Four-Choice tasks.

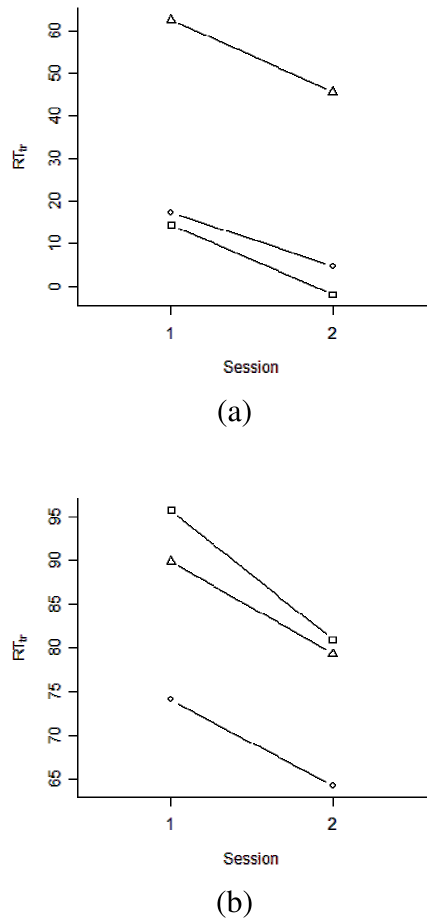


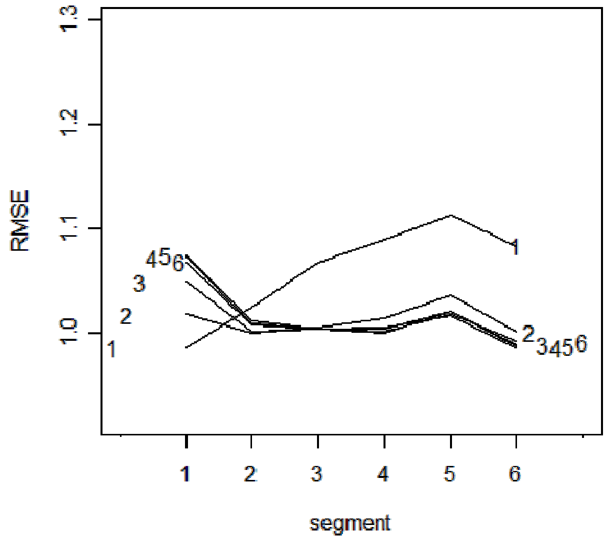
Figure 20. Transformed response time (RT_{tr}) a) means and b) standard deviations across sessions for the No-Choice (☐), Two-Choice (○) and Four-Choice (△) experiments..

Stationarity

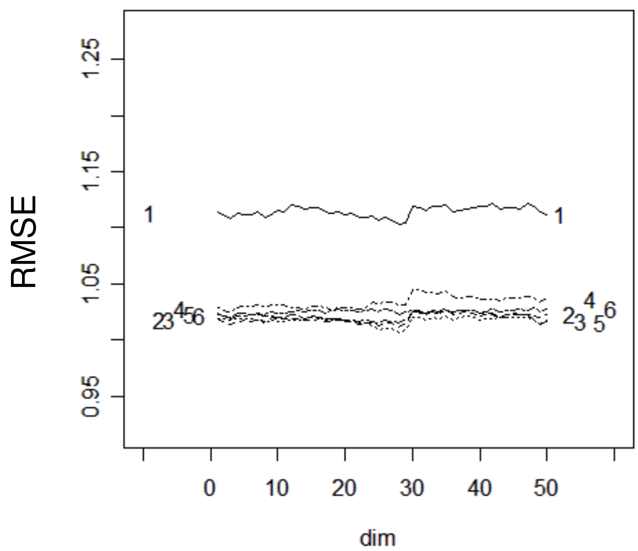
The decreases in mean and standard deviation between sessions offered a first indication of non-stationarity in the data. To examine the time invariance of means and standard deviations across the series with greater detail, they were compared across blocks of trials using a 3 x 20 ANOVA, with Experiment and Block as factors. There was a significant difference in means between blocks, $F(19,855)=35.16$, $p<.01$, but no interaction between blocks and experiments, $F(38,855)=.93$, $p=.58$. There was also a significant difference in block standard deviations, $F(19,855)=34.7$, $p<.01$, with no interaction between experiments and blocks, $F(38,855)=1.38$, $p=.06$.

Initial attempts to resolve the non-stationarity -- for example, by differencing the series, or fitting an exponential function – removed the trend in the block means, but did not remove the trend from the block standard deviations. A Box-Cox transformation was applied to address the trend in the standard deviation. The Box-Cox inverse ($1/x$) transformation successfully removed the change in block means, $F(19,855)=1.37$, $p=.13$, but did not remove the change in block standard deviations, F(19,855)=1.72, p=.02. However, when the first three blocks were removed, the trend in the block standard deviations was also removed, $F(16,675)=.61$, $p=.43$.

While this "weak" stationarity (invariance of means and standard deviations) is satisfactory for linear analysis, nonlinear analysis techniques assume "strict" stationarity, which also requires time invariance of joint probabilities in a series. To determine whether the transformed time series also satisfy the assumption of strict stationarity, the invariance of joint probabilities was examined using Schreiber's (1997) cross-prediction algorithm. Each series was divided into six segments of equal length, with approximately 350 trials in each segment. The cross-prediction algorithm was used to estimate the prediction error for each pair of the six segments.



(a)



(b)

Figure 21. **(a) Cross prediction errors (RMSE) from 6 sequential segments (separate lines) to obtain predictions for all segments, using a 10 dimensional delay 1 embedding, showing that the first segment of the series poorly predicts the remaining 5 segments. (b) The cross-prediction error of each segment averaged over all other segments (separate lines), shown up to 50 dimensions.**

Figure 21a illustrates the relative prediction error for the 2000 trials of a transformed series from the No-Choice estimation experiment. The prediction algorithm used a 10 dimensional, delay 1 embedding, as suggested by Heathcote and Elliott (2005) for analysis of noisy time series. Larger and smaller embedding dimensions produced similar results, as illustrated by Figure 21b, which shows the average prediction error of each segment across all other segments for up to 50 dimensions. Figures 21a and 21b clearly show initial non-stationarity in the series, with the first segment showing poor prediction of the remaining segments. Twenty-four of the forty-eight series showed the same initial non-stationarity, twenty-one of the series showed no evidence of non-stationarity, and three of the series showed intractable non-stationarity. In consideration of the results of the analysis of means and standard deviations, and the cross-prediction analysis, the first 350 trials of the transformed series were removed in order to satisfy the requirement of strict stationary, thereby making the experimental data suitable for linear, nonlinear and fractal analysis techniques. The three intractably non-stationary series were excluded from further analysis.

1/f Noise

Forty-five stationary time series were examined for evidence of 1/*f* noise using a battery of tests: standard Power Spectral Density (PSD) analysis, refined Power Spectral Density analysis (${}^{low}PSD_{we}$), the Disp-SWV method of Eke et al. (2000), and the competitive ARFIMA method of Wagenmakers et al. (2005). The stationary data series, with lengths of approximately 1650 points, were trimmed to 2^n points, a requirement for the analysis techniques that follow. For analysis of the full series, the first 1024 (2^{10})

points of the stationary series were used. For analysis of each session, the first 512 (2^9) stationary points of the each session were used.

To investigate suggestions of spurious reports of 1/f noise resulting from non-stationarity of data, classical and refined spectral analyses were performed on both the first 1024 stationary points and the corresponding non-stationary raw scores.

Examination of the means and standard deviations of the raw data suggested that the participants were learning the task in the first session and had achieved mastery in the second session. The two sessions were therefore examined separately to investigate any change in the dynamical structure of the data series as a result of practice.

The data series were first examined for evidence of 1/*f* noise using standard power spectral density (PSD) analysis. Analysis of the first 1,024 stationary points in each series revealed spectral α values in the range (-.12,.34). Analysis of the raw, non-stationary data for the same 1,024 points yielded spectral α values in the range (-.16, .28). Neither the stationary nor non-stationary alpha values were within the 1/f noise range, α=(0.5,1.5), for any of the experiments.

Although spectral α values for the stationary series (M=.08, SD=.09) were larger than those for the non-stationary series (M=.06, SD=.1), the difference was not significant, $t(92.95)=1.02$, $p=.31$. The non-stationary series were therefore not more likely to be reported as 1/*f* noise than the stationary series.

The data series were then examined for evidence of 1/*f* noise using the refined power spectral density ($^{low}PSD_{we}$) analysis method of Eke et al. (2000). The first 1,024 stationary points of the data series yielded spectral α values in the range (-.42,.68), with two series falling within the 1/*f* noise range: one series from the Two-Choice experiment, and one series from the Four-Choice experiment. The non-stationary series yielded α values in the range (-.32, .5), with only one series from the Four-Choice experiment falling within the 1/*f* noise range. This same series showed evidence of 1/*f* noise using the stationary data. As with standard spectral analysis, failure to resolve the non-stationarity of the raw data would have resulted in fewer rather than more reports of 1/*f* noise than the stationary data using the refined spectral analysis method.

The first and second sessions were examined separately for evidence of 1/*f* noise using the refined power spectral density method. The first session produced spectral α values in the range (-.7, .93); however, only two of the series fell within the 1/*f* noise range: one series from the Two-Choice experiment, and one series from the Four-Choice experiment. These results are similar to those for the full series; however, different series were identified as 1/*f* for the full session and session one analyses. The second session produced spectral α values in the range (-.67, .88). Five of these series were within the 1/*f* noise range: one from the No-Choice experiment, two from the Two-Choice experiment, and two from the Four Choice experiment. Alpha values for the first session (M=.03, SD=.32) were significantly lower than those for the second session (M=.06, SD=.33), $t(47)= 5.95$, $p< .05$. There was therefore more evidence of 1/f noise in the second session, during what is presumed to be automatic performance; however, overall there was no compelling evidence of 1/*f* noise in either session.

Application of Eke et al's (2000) Disp-SWV method to the first 1024 stationary data points produced Hurst (H) coefficients in the range (0.35, .96). Fourteen series fell within the 1/*f* noise region, H=(0.75,1): 1 series from the No-Choice experiment, 3 series from the Two-Choice experiment, and 10 series from the Four-Choice experiment. Application to the non-stationary raw series produced Hurst coefficients in the range (0.43, .97). Nineteen series fell within the 1/*f* noise region. This result is more consistent with reports of non-stationarity producing spurious evidence of 1/*f* noise; however, it is inconsistent with the results of the classical and refined spectral analysis methods.

For the first session, Hurst coefficients from Eke et al's Disp-SWV method were in the range (0.02, .99). Ten series showed evidence of 1/*f* noise: 4 series from the Two-Choice experiment, 6 series from the Four-Choice experiment. For the second session, Hurst coefficients were in the range (.05, .86). Only five series fell within the 1/*f* noise range: 1 series from the No-Choice experiment, and 4 series from the Four-Choice experiment. Only two of the series yielded Hurst coefficients in the 1/*f* noise range in both the first and second sessions, both series from the Four-Choice experiment. The results produced by this method are contrary to those of the refined spectral analysis, where the number of series reported as 1/*f* noise increased in the second session.

Using the competitive ARFIMA modelling framework proposed by Wagenmaker et al. (2004, 2005), 9 ARMA(p,q) models and 9 ARFIMA(p,d,q) models were fitted to the first 1024 scores in each stationary series, with p and q parameters systematically varied from 0 to 2, using the Fracdiff (ARFIMA) library in the R statistical software package. Having fit these models, the determination of long-range dependence involved the determination of whether an ARFIMA(p,d,q) process, which also models long-range dependence, provides a better model of the data (i.e., best fit for minimum complexity) than ARMA(p,q) process, which models only short-range dependence. A series was identified as long-range dependent when the best fitting model was an ARFIMA(p,d,q) process with *d* significantly > 0. A series was identified as 1/*f* noise when the best fitting model was an ARFIMA(p,d,q) process with d > .25.

Wagenmakers et al. (2004) recommend using the Akaike Information Criterion (AIC) to determine the best model; however, Torre, Delignieres and Lemoine (in press) suggest that the Bayes Information Criterion (BIC) provides more accurate results. Both the AIC and BIC were used for comparison. More precisely, AIC and BIC weights (Wagenmakers & Farrell, 2004) were used, rather than raw AIC and BIC values, to determine the best model for each series.

Using AIC weights, an ARFIMA(p,d,q) process with d > 0 provided the best fit for four series, while an ARIMA(p,q) process provided the best fit for the remaining forty-one series. Two of the ARFIMA(p,d,q) series were from Experiment 1, the No-Choice estimation task, with one series from the Fixed ISI condition and one series from the Random ISI condition. The other two ARFIMA(p,d,q) series were from Experiment 2, the Two-Choice estimation task, with both series from the Fixed ISI condition.

Using BIC weights, only two of the forty-five series were identified as an ARFIMA(p,d,q) process with *d* > 0. The remaining forty-three series were best fit by an ARIMA(p,0,q) process, indicating either an i.i.d or short-range dependent process. The two series identified as long-range dependent were from the Four-Choice experiment, with one series each in the Fixed ISI and Random ISI conditions.

However, for all series identified as ARFIMA(p,d,q), *d* < .1. Since the relation between fractional integration *d* values and spectral alpha α values is $\alpha=2d$, these *d* values are equivalent to spectral α < .2, which falls well below the 1/*f* noise region of α=(.5,1.5). Thus, while there is some weak evidence of long-range dependence, there is no evidence of 1/f noise in any series using the competitive ARFIMA(p,d,q) modelling framework.

The ARFIMA framework was then used to examine the data from the first and second session. Specifically, the first 512 points of the first session, and the first 512 points from the second session were examined. The AIC weights identified 20 series as possessing short-range dependence, and 25 series as possessing no dependence between observations. The BIC weights identified all of the first session series as ARMA(0,0), indicating no dependence between observations.

For the second session, both the AIC and BIC weights identified all series as possessing either short-range or no serial correlations. Furthermore, the AIC and BIC weights identified ARMA models with the same p and q values for all but one series in the second session, where the BIC preferred a simpler ARMA(1,0) model to the AIC's ARMA(2,1) model. Therefore, neither the AIC nor BIC identified long-range dependence in general, or $1/f$ noise in particular, in either session.

There did, however, appear to be a difference in the models selected for the No-Choice experiment and the Two- and Four-Choice estimation experiments for each session. This was much more pronounced in the second session. For the No-Choice experiment, AIC and BIC selected ARMA(1,0) or ARMA(1,1) models for all but one series, where they both selected an ARMA(2,1) model. For the Two- and Four-Choice tasks, there was a greater proportion of more complex models. However, there did not appear to be a difference in the number of models of similar complexity between the Two- and Four-Choice experiments, suggesting that the added complexity of the models was related to the addition of the choice component, but was not sensitive to the number of choices.

In summary, the PSD, $^{low}PSD_{we}$, and ARFIMA measures indicated that the majority of the experimental data series contained either no serial correlations or only short-range serial correlations, with long range dependence identified in at most two series in each experiment, and the special case of 1/*f* noise identified in at most one series in each experiment. However, the Disp-SWV method identified 1/*f* noise in fourteen series, including ten series in the Four-Choice experiment (Table 4). Given that these measures have been reported as providing consistent results (e.g. Delignieres et al, 2005), the most parsimonious scenario would be that the PSD, $^{low}PSD_{we}$, and ARFIMA measures were correct and the Disp-SWV produced spurious results.

Table 4. The number of series showing evidence of 1/*f* noise for each experiment using classical spectral density (PSD), refined spectral density ($^{ow}PSD_{we}$), Disp-SWV, ARFIMA using AIC, and ARFIMA using BIC.

	PSD	$^{low}PSD_{we}$	Disp-SWV	ARFIMA (AIC)	ARFIMA(BIC)
No Choice	-	-	1	-	-
Two Choice	-	1	3	-	-
Four Choice	-	1	10	-	-

To assess whether the Disp-SWV method could produce spurious reports of 1/*f* noise from series with no serial correlations or only transient serial correlations, 100 simulated series with no serial correlations were generated, and 100 simulated series with only transient correlations were generated, and analysed using the Disp-SWV method, as well as the ${}^{low}PSD_{we}$ method for comparison. Furthermore, to determine whether the Box-Cox (1/*x*) transform had affected the Disp-SWV results, the series were tested in both their raw form and after applying the Box-Cox transformation.

To create one hundred simulated series similar to the experimental series but with no serial correlations, 1024 independent, identically distributed (i.i.d) values were sampled from a Gaussian distribution with the same mean and standard deviation as the experimental data series. To create one hundred simulated series with transient serial correlations only, the AIC selection criteria was used to determine the best ARMA(p,q) model of the experimental data series, and that model was used to produce one hundred simulated ARMA(p,q) series with the same parameter orders, coefficients and length as the experimental series.

The ${}^{low}PSD_{we}$ analysis of the i.i.d series produced spectral α values in the range (-.4, .4), which fall below the lower bound of the 1/*f* noise range. Disp-SWV analysis of the i.i.d series yielded Hurst coefficients in the range (.3, .79), with four series in the 1/*f* noise range. For the Box-Cox transformed i.i.d series, ${}^{low}PSD_{we}$ analysis produced spectral α values in the range (-.19, .52), with one series falling in the 1/*f* noise range. Disp-SWV analysis of the transformed data produced Hurst coefficients in the range (-.5, .7), which also fall below the 1/*f* noise range. It is evident that the Disp-SWV method can produce a small proportion of spurious reports of 1/*f* noise for series similar to the experimental data but with no serial correlations; however, there were no spurious reports of 1/*f* noise using the Box-Cox (1/*x*) transformation.

The $^{low}PSD_{we}$ analysis of the simulated ARMA series yielded spectral α values in the range (-.2,.67), with five series identified as 1/*f* noise. Disp-SWV analysis of the simulated ARMA series produced Hurst coefficients in the range (.42,.92), with fifty-five series identified as 1/*f* noise. For the Box-Cox transformed data, $^{ow}PSD_{we}$ analysis produced spectral α values in the range (-.6, .53), with only one series identified as 1/*f* noise. Consistent with the raw series, Disp-SWV analysis of the transformed series produced Hurst coefficients in the range (.375, .919), with forty-one series identified as 1/*f* noise. In light of these results, and the original PSD, $^{low}PSD_{we}$, and ARFIMA results, it seems evident that the majority of the experimental data series contained either no or transient serial correlations, which the Disp-SWV method spuriously identified as 1/*f* noise. We therefore defer to the consistent results of the PSD, $^{low}PSD_{we}$, and ARFIMA methods, which indicate that at best only 2 series in each experiment showed evidence of long range dependence, and at best only 1 series in each of the experiments showed evidence of 1/*f* noise.

Nonlinear Analysis

To determine whether any nonlinear structure was present in the data, the 45 stationary time series were tested against the null hypothesis of short-range linear dependence using surrogate series testing. For each of the stationary experimental time series, 99 surrogate series were created using Schreiber and Schmitz's (2000) amplitude adjusted Fourier transform (AAFT) algorithm. The resulting surrogate series possess the linear statistical properties of their respective experimental data series, but do not possess the nonlinear statistical properties. The surrogate series therefore represent the null hypothesis of no nonlinear structure, against which the experimental series can be compared using an appropriate nonlinear statistic.

Before creating the surrogate series, the lengths of the stationary series were truncated to multiples of 2, 3, or 5, to minimise endpoint mismatch. Failure to resolve endpoint mismatch can produce artefacts in the surrogate series that increase the risk of Type II error. This resulted in truncated data series with a typical length of 1536 data points, a loss of approximately 5% of the stationary data.

The surrogate tests used the prediction algorithm of Schreiber and Schmitz (1997) as a measure of nonlinearity. This measure has been found to perform well even with series containing high levels of noise (Heathcote & Elliott, 2005). If the experimental series contains nonlinear structure, the root mean square prediction error (RMSE) of the prediction algorithm for the experimental series should be lower than that of the surrogates, which contain the linear structure but not the nonlinear structure of the experimental series. As with the cross-prediction algorithm used for stationarity testing, a 10 dimensional delay 1 embedding was used for the prediction algorithm.

The choice of 99 linear surrogate series allowed for the creation of 95% confidence intervals for the distribution of root mean square prediction error (RMSE) for the null hypothesis, estimated by the 3^{rd} and 97^{th} percentiles of the surrogate RMSE distribution. This allowed for inferential testing of nonlinear structure. For a nonlinear data series, the RMSE should initially be significantly lower than that of the linear surrogates over several lags, indicating that a linear stochastic model is insufficient to explain the (nonlinear) structure in the series. When these conditions are met, the null hypothesis of no nonlinear structure can be rejected.

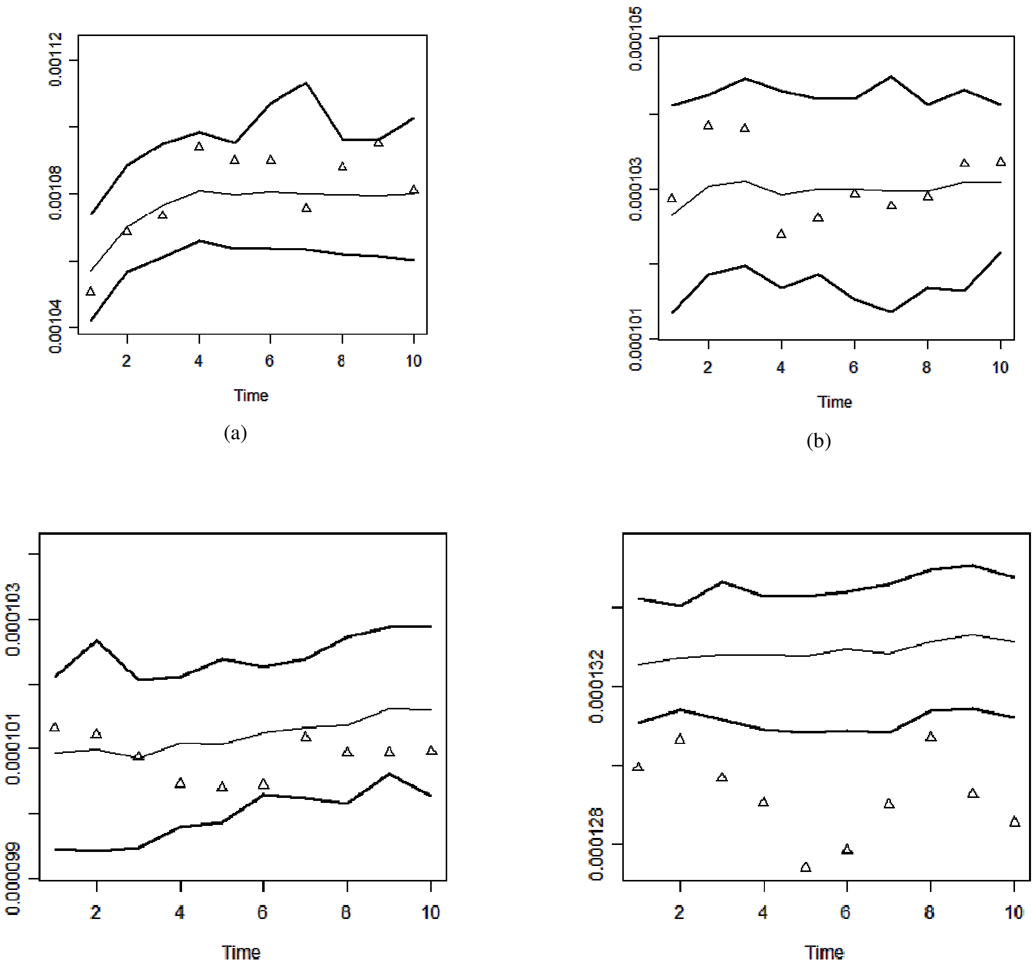
Figure 22 shows the root mean squared prediction error (RMSE) as a function of prediction lag. For the No-Choice estimation experiment, prediction errors for the experimental data were not significantly better than the surrogate data for any of the experimental series (see Figure 22a for an example). Similarly, prediction errors for the Two-Choice experiment also failed to differ significantly from the linear surrogates for any series (see Figure 22b for an example). However, in the Four-Choice experiment, prediction errors for experimental data were significantly better than the surrogate series for one of the experimental series (Figure 22d)[4].

In summary, these results show no compelling evidence of long-range dependence, 1/*f* noise, nor nonlinear dynamics in any of the three estimation tasks. Standard spectral analysis showed no evidence of 1/*f* noise. Refined spectral analysis showed evidence of 1/*f* noise in two series, one each from the Two- and Four-Choice tasks; when the sessions were examined separately, one series each from the Two- and Four-Choice tasks showed evidence of 1/*f* noise in the first session, while one series from the No-Choice estimation task and two series from each of the Two- and Four-Choice tasks showed evidence of 1/*f* noise in the second session. Eke et al's Disp-SWV method provided the most abundant evidence of 1/*f* noise, producing evidence of 1/*f* noise in fourteen of the forty-eight series: one series from the No-Choice estimation task, 3 series from the Two-Choice task, and 10 series from the Four-Choice task. However, it is evident that the results of the Disp-SWV test are spurious. ARFIMA

---

[4] One of the excluded non-stationary series from the Four-Choice experiment also showed evidence of nonlinear structure.

analysis found evidence of long range dependence in four series using the AIC, but found no evidence of 1/*f* noise in any series using the AIC or BIC. Surrogate series testing of experimental data using a nonlinear prediction algorithm found evidence of nonlinear structure in only one series from the Four-Choice task.



(c) (d)

Figure 22. Root mean square prediction error (RMSE) as a function of lag for (a) No Choice estimation task, (b) Two Choice estimation task, (c) Four Choice estimation task. (d) RMSE as a function of lag for the series from the Four Choice task for which the null hypothesis of no nonlinear structure was rejected. Symbols are RMSE for the experimental series. The solid line is the average RMSE of the surrogate series. The bold lines are 95% confidence intervals for the surrogate RMSE.

Discussion

The primary motivation for the current study was to investigate possible relationships between nonlinearity, 1/*f* noise, and task demands in the context of an interval estimation paradigm, as previously investigated by Elliott (2003). In the previous study, no evidence of nonlinearity was found in data from the Slow Pace experiment, some evidence of nonlinearity was found in the data from the Fast Pace experiment, and no evidence of 1/f noise was found in either experiment. Although the evidence was not compelling, it did offer the possibility of a relationship between nonlinear dynamics and task pace similar to that seen in the Choice RT task of Kelly et al. (2001), where a positive relationship was found between the pace of the task and the number of series showing evidence of nonlinear dynamics. However, Elliott found no evidence of 1/f noise at any pace. Elliott suggested that only a small number of series showed evidence of nonlinear dynamics in the Fast Pace experiment because the task was not demanding enough to elicit such dynamics from all participants. A similar pattern was seen in experiment two of Kelly et al. (2001), with no evidence of nonlinear dynamics in the self-paced condition, evidence of nonlinear dynamics in some series from the slow force-pace condition, and evidence of nonlinear dynamics in all series from the fast force-pace condition. Elliott (2003) speculated that further increasing the task demands might yield a greater number of series with evidence of nonlinear dynamics in his paradigm.

The current study, therefore, modified the task from Elliott (2003) to increase the task demands in two ways. First, in experiment 1, the pace of the fast task in Elliott (2003) was increased by effectively halving the ISI and temporal estimation interval (No-Choice task). Second, in experiments 2 and 3, the fast pace task in Elliott (2003) was made more complex by adding a choice component, with two and four choices (Two- and Four-Choice experiments, respectively); this also made experiments 2 and 3 more analogous to Kelly et al. (2001). It was expected that the addition of a choice component would yield a substantial number of data series with evidence of nonlinear dynamics, but not 1/f noise.

However, comparison of the proportion of missed responses, errors, accuracy, and precision in Elliott (2003) and experiments 1-3 of the current study suggest that these modifications may not have substantially increased the task demands. In Elliott (2003), the miss rates for the Slow and Fast Pace experiments were .25% and .95%, respectively. In the current study, the error rates for the No-Choice, Two-Choice and Four-Choice experiments were 1.55%, 1.36% and 1.88%, respectively. Although the miss rate for the Four-Choice experiment in the current study is twice that of the Fast pace experiment in Elliott (2003), it is not surprising that an increase in missed responses from 1% to 2% is not associated with a dramatic change in sequential structure.

In Elliott (2003), participants in the Fast pace experiment under-estimated the target interval by 10ms, while participants in the No- and Two-Choice experiments of the current study over-estimated the target interval by 10ms. Standard deviations were also similar for Elliott (2003) and the present study, starting at approximately 110ms and dropping to approximately 80ms in the second session. This seems to suggest that task demands in the current study, at least as quantified by the miss rate, were the same as, if not lower than, those of Elliott (2003).

The small difference in missed responses and errors may explain the failure to achieve a greater yield of data series with nonlinear dynamics. If a relationship exists between task demands and nonlinear dynamics, and the experiments in the current study are no more demanding than those in Elliott (2003), then naturally we would not expect to see a greater number of series with evidence of nonlinear dynamics in the current study. Of course, this assumes that the relationship between task demands and nonlinear dynamics is proportionate and the manipulation of task demands sufficient.

Alternately, these results may highlight the inadequacy of using the proportion of missed responses as a measure of task demands. It seems counter-intuitive to believe that a task is more demanding than the same task performed at twice the pace. It is more plausible that the miss rate – the number of missed responses -- reflects the degree of engagement. Participants who are not as fully engaged in the task are more likely to miss responses. In this interpretation, the Two- Choice experiment, with the highest miss rate, was the least engaging, while the No- and Four- Choice task were equally engaging. However, with miss rates differing by only .5% between the Two-Choice and Four-Choice experiments, the difference may be statistically significant, but in practical terms, missing .5% more responses suggests a negligible difference in the degree of engagement between experiments. The difference in the number of incorrect responses between the Two-Choice and Four-Choice experiments, on the other hand, reflects a more accurate measure of task demands. With approximately 9% of responses incorrect in the Two-Choice experiment, and approximately 22% of responses incorrect in the Four-Choice experiment, the Four-Choice experiment is clearly more difficult than the Two-Choice experiment. These results suggest that the experiments were equally engaging, but that the Four-Choice experiment was more difficult.

If we accept that task demands have increased, it remains to explain why a greater proportion of series do not show evidence of nonlinear dynamics. The most parsimonious explanation is that there is no relationship between task demands and nonlinear dynamics in these tasks. It is possible that nonlinear dynamics are the result of some unknown factor, or combination of factors, shared by less than 2% of participants in Elliott (2003) and the current study. It is also possible, and more likely, that the nonlinear dynamics reported here and in Elliott (2003) are spurious. The proportion of series is more akin to a type I error rate than a genuine effect. Another explanation, one of the least parsimonious given the available data, is that already offered by Elliott (2003): there is a relationship between nonlinear dynamics and pace, but the pace of the tasks are not yet sufficiently fast enough to elicit nonlinear dynamical structure from the majority of participants. Given the failure of the current study to support this hypothesis, it would be difficult to justify the expectation of more fruitful results if the pace was increased yet again. Furthermore, the pace of the current task is such that it could only be increased a little more before the task becomes a reaction task, and the choice tasks become similar to the Choice RT tasks in Kelly (2001).

However, regarding the possibility that the demands are not yet sufficient to elicit nonlinear dynamics, a comparison of the miss levels reported by Kelly et al. (2001) and those in the current study, and the proportion of series with nonlinear dynamics, suggest that the results in the current study are at least consistent with a relationship between error levels and the proportion of series showing evidence of nonlinear dynamics. Kelly et al. found that the error level increased with pace, with 1.4% missed in the slow-pace experiment, in which 25% of series were identified as nonlinear, and 6.9% missed in the fast-pace experiment, in which 100% of series were identified as nonlinear. That is, the number of missed responses increased with the number of series showing evidence of nonlinear dynamics. In the current study, the miss rate for the Two Choice estimation experiment was .45%, one-quarter of the miss rate of Kelly et al.'s Slow Pace experiment. If, for the sake of argument, it is assumed that there is a correlation between the error level and the proportion of series showing evidence of nonlinear dynamics, then experiment 2, with an error level of .45%, would expect to contain .06% or one series showing evidence of nonlinear dynamics, while in experiment 3, with an error level of 1.4%, would expect to contain 25% or four series showing evidence of nonlinear dynamics. In reality, there were no series with nonlinear dynamics in experiment 2, and only one series with nonlinear dynamics in experiment 3. Although the suggestion of such a simple relationship is attractive, the current study does not support it.

There is also a methodological difference between the task in Kelly et al. (2001) and the task in the current study that might explain the failure to achieve similar results. Kelly et al. tailored their task to suit individual differences, by setting each participant's ISI as a function of that participant's RT in a self-paced condition. The current study did not adjust ISI for individual participants, but instead used a single ISI for all participants in each experiment. Had we adjusted each participant's ISI, the task may have been demanding enough for each participant to produce data series containing nonlinear structure. By this reasoning, we found evidence of nonlinear dynamics in only one series because the chosen ISI was incidentally appropriate for one of sixteen participants.

Another less parsimonious explanation is that a number of series in the current experiments do possess nonlinear dynamical structure, but this structure is masked by large amounts of additive noise. In Kelly's (2001) study of speeded choice RT, she identified nonlinear dynamics in 100% of data series from participants in the fastest pace. Furthermore, Kelly noted that many of the reconstructed attractors for participants in her choice RT experiments resembled the Lorenz attractor, with an unstable saddle and two adjacent unstable cycles. The task in Elliott (2003) and experiments 1-3 of the current study attempted to replicate the dynamics found by Kelly, but used larger response times and response boundaries. If the Lorenz-like attractor reported by Kelly resulted from the entrainment of response times via the interaction between response boundaries, then a similar attractor may emerge from a similar entrainment of response times via the interaction between response boundaries. However, this was not the case. In Elliott (2003) and the current study, at most two series showed evidence of nonlinear dynamics. However, it is possible that Elliott (2003) and the current study, by increasing the interval between response boundaries, dampened the interaction between the boundaries, and thereby dampened, or removed, any nonlinear dynamics.

Heathcote and Elliott (2005), using the same nonlinear analysis techniques that were employed in the current study and in Elliott (2003), identified the nonlinear dynamics of the Lorenz system with up to 80% additive noise. It is interesting to note that the results from the nonlinear analysis of the data in the current study present similarities to those of nonlinear analysis of the noisy Lorenz data in Heathcote and Elliott. For the two series identified as containing nonlinear structure, the log-log power versus frequency plots and the nonlinear prediction error plots resemble the power spectra and nonlinear prediction error plots for the Lorenz series with 80% additive noise. Of course, this is necessary but not sufficient evidence for a Lorenz-like attractor with large amounts of noise. It is nonetheless appealing to consider that a greater number of series may have nonlinear structure which is masked by additive white noise, rather than suggest that nonlinear dynamics emerges in only 2% of series. Is there more gold buried beneath the surface, or has the current study merely discovered fools' gold?

The rare occurrence of 1/*f* noise presents a similar situation. If we disregard the results of the Disp-SWV analysis, 1/*f* noise was identified in only one series (6%) from each of the Choice experiments, and only when using the $^{low}PSD_{we}$ measure. Had the current study relied on standard spectral analysis, or ARFIMA analysis, there would be no evidence of 1/*f* noise in any of the data series. The most parsimonious interpretation of these results is to assume that the identificaitons of two series as containing 1/*f* noise using the $^{low}PSD_{we}$ analysis are spurious, most likely false positives, and the data series contain no 1/*f* noise. Yet, even allowing for the results of the $^{low}PSD_{we}$ analysis, the results fail to provide any compelling evidence of 1/*f* noise. Pragmatically, the discovery of 1/*f* noise in 6% of data series serves no descriptive or prescriptive value. Certainly, in terms of Gilden's claims regarding the ubiquity of 1/*f* noise, the results of Elliott (2003) and the current study show 1/*f* noise not to be pervasive.

Comparison of the first and second sessions using fractal analysis techniques also failed to produce definitive results. There is a clear warm-up effect in the first session, with the local mean and standard deviation decreasing exponentially with time. In the second session, the warm-up effect is present only for the initial blocks. Local means and standard deviations were minimal and invariant for the rest of the second session. If learning is associated with controlled processing, and mastery is associated with automated processing, then the first session should reflect controlled processing, while the second session should reflect automatic processing. Following Heiby et al. (2003), correlated noise would be expected to dominate in the first session, while white noise would be expected to dominate in the second session. However, comparison of the fractal analysis results for the two sessions shows, if anything, the opposite effect. Those measures which show evidence of 1/*f* noise show more series containing 1/*f* noise in the second session that the first.

The discrepancy in results from the four measures of 1/*f* noise highlights the difficulties facing researchers attempting to identify 1/*f* noise in psychological data. It also highlights the utility, as recommended by Rangarajan and Ding (2000), of performing a battery of tests rather than relying on a single measure. Quite different results would have been obtained had the current study relied on the Disp-SWV measure. Without the divergent results of the other three measures, the results of the Disp-SWV analysis would never have been questioned, and the subsequent simulation testing would not have shown the Disp-SWV method could produce spurious reports of 1/*f* noise in up to 50% of series containing only transient dependencies. Whether this fault can be generalised, or is limited to the context of the current study, is beyond the scope of this thesis.

Contrary to expectations, non-stationary data appeared to make identification of 1/*f* noise in the data series less likely. For the standard spectral analysis, alpha values for the non-stationary data were the same as the stationary data. Refined spectral analysis produced similar results to standard spectral analysis, with alpha values for the non-stationary data lower than those for the stationary data. The Disp-SWV method produced results in the expected direction, with a larger number of series being identified as 1/f noise using non-stationary data; however, because the Disp-SWV method produced possibly spurious results, these results may simply indicate that the Box-Cox transformation dampened the transient correlations in the series. Although the original data was non-stationary, this was due to a very slight quadratic trend in the mean and standard deviation. It did not approach the degree of non-stationarity which Pressing et al. (1997) suggested as an explanation for the apparent 1/f noise reported by Gilden et al. (1995). It is, therefore, not surprising that the results of spectral analysis of the original, non-stationary data did not differ significantly from the results for the stationary data, which showed no compelling evidence of 1/f noise.

The Role of Feedback

It is possible to view the task in the experiments reported here as a form of synchronous tapping. When the participant responds, the colour gradient of the stimulus changes colour to reflect the accuracy of their response. This is equivalent to providing explicit feedback about the asynchrony between the response and a metronome. This could also explain the lack of structure in the responses, as fluctuations in asynchronies for metronome tapping tasks have been found to be independent (e.g., Musha et al., 1985).

Subsequent to the experiments performed here, and in order to attempt to explain the striking lack of any dynamic structure, the author found papers by Spray (1981) and Spray and Newell (1986) examining the effect of feedback on temporal and spatial estimation tasks. They identified sequential structure only in those conditions where feedback about performance was withheld from participants. Spray performed ARIMA analysis on the time series data obtained from two earlier studies by Diggles (1977) and Turpin (1977). Diggles presented participants with a line of lights which were illuminated in succession to simulate movement. Participants were required to press a response key when they anticipated the arrival of the light stimulus at the final position. On each trial participants were provided with feedback about the accuracy of their response. Note that the task presented in Elliott (2003) and in the current study is functionally equivalent to that presented by Diggles, in that it required participants to anticipate the arrival of illuminated pixels at their destination rather than illuminated lights. Turpin required participants to displace a shoulder-height metal bar by 12 inches in 180ms. On each trial, their precision (in ms, greater or lesser than 180ms) was provided. Both Diggles and Turpin also tested control groups which received no feedback about their performance. Spray examined group-averaged data and reported that the responses of participants receiving feedback in both the Diggles and Turpin studies were best fit by a white noise model, while responses of participants who did not receive feedback were best fit by a non-stationary random walk model.

Spray and Newell (1986) performed the same analysis with data from a spatial estimation experiment by Newell (1974), similar to that of Turpin (1978). In the Newell (1974) study, participants were required to displace a slider along a track by 10cm in 150ms. Twelve groups received various levels of trial-by-trial feedback about their performance. Spray and Newell (1986) focused their analysis on only five of the groups: feedback, no feedback, feedback stopped after 17 trials, feedback stopped after 32 trials, and feedback stopped after 52 trials. They first examined group-averaged data and found that for the feedback groups, including the feedback-withdrawal groups up to the point where feedback was withdrawn, the responses were best fit by a white noise model. For the no feedback group, the responses were a random walk process. For the no-feedback trials of the feedback-withdrawal groups, the responses were random walks when the feedback was withdrawn after 17 and 32 trials (although it was closer to white noise for the 32 trials group), while the responses were white noise when the feedback was withdrawn after 52 trials. They then examined individual time series and found that all of the feedback series were white noise, consistent with the group mean series. However, for the no-feedback group, there was a variety of models amongst participants: nine series were consistent with the group mean series, seven were white noise series, and four were various ARIMA models. Spray and Newell concluded that in the absence of feedback, individuals use a variety of processes to learn the criterion for the task; to paraphrase Charles Dickens, all subjects with feedback perform alike, but each subject without feedback performs in their own way.

The hypothesis that it may be the presence of feedback about the accuracy of responses in the task examined in this thesis and in Elliott (2003) that is responsible for the absence of 1/f noise, and possibly also the absence of nonlinear dynamical structure, is supported by the observation that previous studies which reported 1/f noise and nonlinear dynamics, did not provide feedback about accuracy of responses to participants (e.g. Aks, Zelinsky, & Sprott, 2002; Boker, 1996; Chen, Ding, Kelso, 1997; Clayton & Frey, 1997; Delignieres et al, 2001; Gilden et al., 1995; Gilden, 1997; Kelly et al., 2001; Madison, 2001; Musha et al. 1985; Pressing, 1999; Ward, 2002).

A simple modification of the current task would eliminate feedback and determine whether it was responsible for the lack of 1/f noise. A practice block of trials would be provided using the same methodology as the current task, followed by experimental blocks, in which the stimulus is not shown, and either no feedback is provided, or simple feedback (such as 'hit' or 'miss') is provided. The task would then be more analogous to the interval production tasks (e.g. Gilden & Thornton, 1995; Musha et al., 1985; Wagenmakers et al., 2004; Yamada et al., 1993). If 1/*f* noise was discovered, it would support the notion that 1/f noise is pervasive only in situations where feedback is absent.

Whether the failure to identify 1/f noise in the experimental data can be attributed to methodological differences between the current study and studies which did identify 1/f noise, these results do not support the assertion by Gilden and colleagues that 1/f noise is ubiquitous in time series data from psychological studies. It could be argued that the failure to identify 1/f noise can be attributed to the current study using a novel interval estimation task, rather than standard interval estimation tasks where 1/f noise has been identified. However, if 1/f noise is a ubiquitous phenomena, as Gilden and colleagues believe, it is reasonable to expect to find evidence of 1/f noise in more than a small number of standard laboratory tasks. In biology, physics, geology, economics, and other fields where 1/f noise has been identified, evidence of 1/f noise has been found “in the wild”. It is not merely reported in a small number of isolated laboratory experiments with carefully controlled parameters. It is found in the landscape and ebbing of the tides. 1/f noise cannot be said to be ubiquitous in this sense. However, if a connection between 1/f noise and feedback exists, 1/f noise in psychological data could be said to be a natural phenomena, rather than an ubiquitous one. From this perspective, feedback could be seen to disrupt the natural performance of the system, enforcing an artificial performance regime that adheres to the criteria determined by the feedback. 1/f noise in psychological data may, therefore, be limited to contexts where natural performance is not disrupted by external stimuli (e.g. a metronome).

It is apparent that any sequential structure in the data from the present experiments can be best characterised as white noise or at best short-range dependent, low-order ARIMA(p,q) processes. Although the data offer glimpses of 1/f noise, and a single glimpse of nonlinear structure, knowing that one or two series may be characterised by such complex sequential structure serves no practical utility. Whether these few series present a genuine occurrence of these complex phenomena, with some unknown factor possessed by 2% of participants responsible for their appearance, or whether they are simply false positives, pragmatically, these data are inconsistent with a nonlinear RT control strategy, and demonstrate that 1/f noise is not a pervasive phenomena.

References

Aks, D., & Sprott, J. C. (2003). The Role of Depth and 1/f Dynamics in Perceiving Reversible Figures. *Nonlinear Dynamics, Psychology, and Life Sciences*, 7(2), 161-180.

Arons, L., & Irwin, F. W. (1932). Equal weights and psychophysical judgements. *Journal of Experimental Psychology*, 15, 733-756.

Auyang, S. Y. (2000). *Mind in Everyday Life and Cognitive Science*. Cambridge, MA: MIT Press.

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

Barton, S. (1994). Chaos, Self-Organisation, and Psychology. *American Psychologist*, 49(1), 5-14.

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

Beran, J. and Mazzola, G. (2001) Musical composition and performance – statistical decomposition and interpretation. *Student*, 4(1), 13-42.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 13, 90-102.

Boker, S. M. (1996). Linear and nonlinear dynamical systems data analytic techniques and an application to developmental data. *Unpublished PhD thesis,* University of Virginia, Virginia.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and Control.* San Francisco: Holden Day.

Brewer, N., & Smith, G. (1989). Developmental Changes in Processing Speed: Influence of Speed-Accuracy Regulation. *Journal of Experimental Psychology: General*, 118(3), 298-310.

Busey, T., & Townsend, J. (2001). Independent Sampling vs Interitem Dependencies in Whole Report Processing: Contributions of Processing Architecture and Variable Attention. *Journal of Mathematical Psychology*, 45, 283-323.

Caccia, D. C., Percival, D., Cannon, M. J., Raymond, G., & Bassingthwaighte, J. B. (1997). Analyzing exact fractal time series: evaluating dispersional analysis and rescaled range methods. *Physica A,* 264(609-632).

Campbell, M. J. and Jones, B.W (1972). Cyclic changes in insulin needs of an unstable diabetic. *Science*, 177, 889-891.

Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology,* 27(2), 198–208.

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

Casdagli, M. (1997). Recurrence plots revisited. *Physica D*, 108, 12–44.

Chen, Y., Ding, M., & Kelso, J. (1997). Long Memory Processes (1/f Type) in Human Coordination. *Physical Review Letters*, 79(22), 4503-4504.

Chen, Y., Repp, B. R., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, 21, 515-532.

Choi, M.Y, and Lee, H.Y. (1995), Traffic flow and 1/f fluctuations, *Physical Review E*, 52(6), 5979-5984.

Chow, C., & Collins, J. (1995). Pinned polymer model of posture control. *Physical Review E*, 52, 907–912.

Clayton, K., & Frey, B. (1996). Inter- and Intra-Trial Dynamics of Memory and Choice. In W. Sullis & A. Combs (Eds.), *Nonlinear Dynamics in Human Behaviour* (Vol. 5). Singapore: World Scientific.

Delignières, D., Torre, K., & Lemoine, L. (2005). Methodological Issues in the Application of Monofractal Analyses in Psychological and Behavioral Research. *Nonlinear Dynamics in Psychology and Life Sciences*, 10(4), 435.

Donders, E C. (1969). On the speed of mental processes. In W. G. Koster (Ed. and Trans.), *Attention and performance H* (pp. 412-431). Amsterdam: North Holland. (Original work published 1868)

Diggles, V. A. (1977). The applicability of Adams' closed loop theory of motor learning to perceptual learning. *Unpublished thesis*, California State Polytechnic University, California.

Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement,* 23, R1-R38.

Eke, A. H., Bassingthwaighte, J.B., Raymond, G. M., Percival, D. B., Cannon, M., Balla, I., & Ikrényi, C. (2000). Physiological time series: distinguishing fractal noises from motions. *European Journal of Physiology*, 439, 403-415.

Elliott, D. (2003). Low-dimensional nonlinear dynamics in temporal estimation. *Unpublished thesis,* University of Newcastle.

Engbert, R., & Kliegl, R. (2004). Microsaccades Keep The Eyes Balance During Fixation. *Psychological Science*, 15(6), 431-436.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews: Neuroscience,* 4, 1-9.

Fernberger, S. W. (1920). Interdependence of judgements within the series for the method of constant stimuli. *Journal of Experimental Psychology*, 3, 126-150.

Fraedrich, K., Blender, R. (2003), "Scaling of atmosphere and ocean temperature correlations in observations and climate models", Physical Review Letters, 90:108501.

Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33, 1134–1140.

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

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

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

Gleick, J. (1987). *Chaos: Making a New Science*. London: Cardinal.

Gottman, J. M., Murray, J. D., Swanson, C. C., Tyson, R., & Swanson, K. R. (2005). *The mathematics of marriage: dynamic nonlinear models*. Cambridge, MA: MIT Press.

Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51, 347-356.

Hale, D. J. (1969). Repetition and probability effects in a serial choice reaction task. *Acta Psychologica*, 29, 163-171.

Handel, P. H., & Chung, A. L. (1993). *Noise in physical systems and 1/f fluctuations*. New York, NY: American Institute of Physics.

Heath, R. (2000). *Nonlinear Dynamics: techniques and applications in Psychology*. London: Lawrence Erlbaum Associates.

Heathcote, A., & Elliott, D. (2006). Nonlinear dynamical analysis of noisy time series. *Nonlinear Dynamics in Psychology and Life Sciences*, 9, 399-434.

Heiby, E. M., Pagano, I. S., Blaine, D. D., Nelson, K., & Heath, R. A. (2003). Modeling unipolar depression as a chaotic process. *Psychological Assessment*, 15(3), 426-434.

Hoyle, R. H., & Gohlson, B. (1968). Choice reaction times with equally and unequally probably alternatives. *Journal of Experimental Psychology*, 78, 95-98.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 45, 188-196.

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, 41, 291-297.

Kantz, H., & Schreiber, T. (1997). *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press.

Kaplan, D., & Glass, L. (1995). *Understanding Nonlinear Dynamics*. New York: Springer-Verlag.

Karagiannis, T., Faloutsos, M., & Riedi, R. H. (2002). Long range dependence: now you see it, now you don't? Paper presented at the Proc. IEEE Global Telecommunications Conf. Global Internet Symp.

Kelly, A. (2001). Non-linear determinism in human information processing: evidence for self-organised response strategies under time-constrained task conditions. *Unpublished Ph.D thesis*, University of Newcastle, Newcastle.

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

Kelso, J. (1995). *Dynamic Patterns: Self-Organisation of Brain and Behaviour*. Cambridge, MA: MIT Press.

Kennel, M. B., Brown, R., & Abarbanel, H. D. I. (1992). Determining embedding dimension for phasespace reconstruction using a geometrical construction. *Physical Review A*, 45, 3403–3411.

Kirby, N. H. (1975). Sequential effects in an eight choice serial reaction time task. *Acta Psychologica*, 39(39), 205-216.

Kornblum, S. (1967). Choice reaction time for repetitions and nonrepetitions: a re-examination of the information hypothesis. *Acta Psychologica*, 27, 178-187.

Kornblum, S. (1968). Serial-choice reaction time: inadequacies of the information hypothesis. *Science*, 159, 432-434.

Kornblum, S. (1969). Sequential determinants of information processing in serial and discrete choice reaction time. *Psychological Review*, 76, 113-131.

Kornblum, S. (1973). Sequential effects in choice RT. In S. Kornblum (Ed.), *Attention & Performance IV* (pp. 259-288). New York: Academic Press.

Kulessa, B., Srokowski, T., Drozdz, S. (2003), Long-time autocorrelation function of ECG signal for healthy versus diseased human heart, *Acta Physica Polonica B*, 34(1), 3-15.

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

Leonard, J. S., Newman, R. C., & Carpenter, A. (1966). On the handling of heavy bias in self paced tasks. *Quarterly Journal of Experimental Psychology*, 18, 130-141.

Liu, M. (2000), "Modeling long memory in stock market volatility", *Journal of Econometrics*, 99(1):139-171.

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of Atmospheric Science*, 26, 636.

Luce, R. D. (1995). Four tensions concerning mathematical modelling in psychology. *Annual Review of Psychology*, 46, 1-26.

Madison, G. (2001). Functional Modelling of the Human Timing Mechanism. *Unpublished Doctoral Dissertation*, Uppsala University, Uppsala, Sweden.

Mitra, S., Riley, M., & Turvey, M. (1997). Chaos in human rhythmic movement. *Journal of Motor Behavior*, 29, 195-198.

Musha, T., Katsurai, K., & Teramachi, Y. (1985). Fluctuations of Human Tapping Intervals. *IEEE Transactions of Biomedical Engineering*, BME-32(8), 578-581.

Newell, K., & Slifkin, A. (1998). The nature of movement variability. In J. Piek (Ed.), *Motor behaviour and human skill: a multidisciplinary Approach*. Illinois: Human Kinetics.

Peak, D., & Frame, M. (1994). *Chaos Under Control: The Art and Science of Complexity*. New York: W. H. Freeman and Company.

Pressing, J. (1999). The referential dynamics of cognition and action. *Psychological Review*, 106, 714-747.

Pressing, J. (1999). Sources for 1/f noise effects in human cognition and performance. *Paper presented at the Proceedings of the 4th Conference of the Australasian Cognitive Science Society*., Newcastle, Australia.

Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics*, 76, 339-347.

Preston, M. (1936). Contrast effects and the psychophysical judgements. American *Journal of Psychology*, 48, 625-631.

Rabbitt, P. (1965). Response facilitation on repetition of a limb movement. British *Journal of Psychology*., 56, 303-304.

Rabbitt, P. (1968). Repetition effects and signal classification strategies in serial choice-response tasks. *Quarterly Journal of Experimental Psychology*, 20, 232-239.

Rabbitt, P. (1992). Many happy repetitions: A celebration of the "Bertelson Repetition Effect" 1961-1991. In J. Alegria & D. Holender (Eds.), *Analytic approaches to human cognition.* (pp. 313-330).

Rabbitt, P., & Vyas, S. M. (1970). An elementary preliminary taxonomy for some errors in laboratory choice RT tasks. *Acta Psychologica*, 33, 56-76.

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

Remington, R. J. (1971). Analysis of sequential effects for a four-choice reaction time experiment. *Journal of Psychology*, 77, 1211-1216.

Riley, M., & Turvey, M. (2002). Variability and Determinism in Motor Behaviour. *Journal of Motor Behavior*, 34(2), 99-125.

Ryabov, V. B., Stepanov, A.V., Usik, P.V., Vavriv, D.M., Vinogradov, V.V, Yurovsky, Y.U. (1997), "From chaotic to 1/f processes in solar mcw-bursts", *Astronomy & Astrophysics*, 324:750-762.

Saiki, T., Kitagawa, Y., and Hayashi, A. (1999), "Fluctuations of character centroid intervals in laterally written Japanese sentences", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E82A(3), 520-526.

Schreiber, T. (1999). Interdisciplinary application of nonlinear time series methods. *Physics Reports*, 308, 1–64.

Schreiber, T., & Schmitz, A. (1997). Discrimination power of measures for nonlinearity in a time series. *Physical Review E*, 55, 5443–5447.

Schreiber, T., & Schmitz, A. (2000). Surrogate time series. *Physica D*, 142, 346–382.

Schvaneveldt, R. W., & Chase, W. G. (1969). Sequential effects in choice reaction time. *Journal of Experimental Psychology*, 80, 1-8.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. . *Psychological Review*, 84(2), 127-190.

Shuster, H. G. (1996). *Deterministic chaos: an introduction.* Cambridge: Weinheim.

Smith, M. (1968). The repetition effect and short term memory. *Journal of Experimental Psychology*, 77, 435-439.

Spray, J. A. (1981). Time series modeling of motor task responses. *Unpublished Doctoral Dissertation*, University of Wisconsin, Madison.

Spray, J. A., & Newell, K. M. (1986). TIME SERIES ANALYSIS OF MOTOR LEARNING: KR VERSUS N0-KR. *Human Movement Science*, 5, 59-74.

Stewart, I. (1989). *Does God Play Dice?: The Mathematics of Chaos*. Oxford: Basil Blackwell.

Sugihara, G., & May, R. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344, 734–741.

Takens, F. (1981). *Detecting strange attractors in turbulence.* (Vol. 898). New York: Springer.

Thagard, P. (2005). *Mind: Introduction to cognitive science.* Cambridge, MA: MIT Press.

Theiler, J. (1990). Estimating fractal dimension. *Journal of the Optical Society of America A*, 7, 1055–1073.

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

Thelen, E., & Smith, L. (1995). *A Dynamic Systems Approach to the Development of Cognition and Action.* Reading, MA: MIT Press.

Thornton, T., & Gilden, D. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review.*, 12, 409-441.

Townsend, J., Hu, G., & Kadlec, H. (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. *Perception & Psychophysics*, 43, 575-591.

Tsonis, A. A. (1992). *Chaos: From Theory to Applications*. New York: Plenum Press.

Turpin, B. A. M. (1978). Response recognition of a rapid, discrete, gross movement: implications of Adams' closed-loop theory of motor learning. Unpublished Masters thesis, Purdue University.

Van Orden, G. C., Moreno, M. A., & Holden, J. G. (2003). A proper metaphysics for cognitive performance. *Nonlinear Dynamics, Psychology, and Life Sciences*, 7(1), 49-60.

Veres, A., Kenesi, Z., Molnar, S., Vattay, G. (2000), On the propagation of long-range dependence in the internet. *Computer Communication Review*, 30(4), 243-254.

Wagenmakers, E.-J., & Farrell, S. (2004). A note on model selection using the AIC. *Psychonomic Bulletin & Review*. 11(1), 192-196

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/fa noise in human cognition. *Psychonomic Bulletin & Review*., 11, 579-615.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General*, 134, 108-116.

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

Ward, L. M., & Richard, C. M. (2001). 1/*f* noise and decision complexity. Unpublished manuscript, University of British Columbia.

Wertheimer, M. (1953). An investigation of the "randomness" of threshold measurements. *Journal of Experimental Psychology*, 45, 294-303.

Wing, A. M., & Kristofferson, A. B. (1973). The timing of inter-response intervals. *Perception and Psychophysics*, 13, 455-460.

Wolf, M. (1997), 1/*f* noise in the distribution of prime numbers, *Physica A*, 241, 493-499.

Yamada, M. (1996). Temporal Control Mechanism in Equaled Interval Timing. *Applied Human Science*, 15(3), 105-110.

Yamada, M., Imura, K., & Takabatake, K. (1993). Temporal fluctuation in equaled time interval tapping by pianists as a function of their skill. *Paper presented at the Proceeding of the international symposium on design of amenity, Fukuoka.*

Appendix A

Details of Fractal Analysis Measures

Power Spectral Density Analysis (PSD)

Power spectral density analysis (aka the periodogram method, or Fourier analysis) has been the preferred method for detecting long range dependencies reported in the psychophysical literature. Fourier analysis approximates a time series as the sum of sinusoidal waves of varying frequencies, amplitudes, and phases (Tsonis, 1992). A power spectrum is the bi-logarithmic plot of power (the square of amplitude) versus frequency, with the scaling exponent α estimated by the slope of a linear regression of log-power versus log-frequency. The fluctuations in the experimental series are then characterised according to the value of α.

Refined Power Spectral Density Analysis ($^{low}PSD_{we}$)

A refined method, $^{low}PSD_{we}$, proposed by Eke et al. (2000) adds a number of pre-processing steps. For each value in the series, the mean is subtracted, and a parabolic window is applied.

Detrending is performed by removing the linear regression of the first and last values of the series, and high frequency power values are excluded for frequencies higher than 1/8 of the maximum frequency.

Rescaled Range Analysis (R/S)

Rescaled range analysis (Hurst, 1965) divides the experimental series into non-overlapping segments of length *n.* Each segment is integrated, and the range *R* of each segment is calculated and normalised by dividing it by the standard deviation of the segment. This is repeated for increasing sequence lengths *n*. In the *rescaled R/S* method (Caccia et al, 1997), each sequence is detrended by removing the linear fit between the first and last points before calculating the range *R*. The slope of the regression of the plot of log-*R* versus log-*n* provides the scaling exponent.

Detrended Fluctuation Analysis (DFA)

Detrended Fluctuation Analysis (Peng, Mietus, Hausdorff, Havlin, Stanley, and Goldberger, 1993) integrates the deviation of each value in the series from the mean, then divides the integrated series into non-overlapping segments of length *n*. Each segment is detrended by removing the best linear fit, and the fluctuation *F* of each segment is calculated using

This process is repeated for increasing values of *n*. The linear regression of a bi-logarithmic plot of *F* versus *n* provides an estimate of the scaling exponent.

Dispersional Analysis (Disp)

Dispersional analysis (Bassingthwaighte, 1988) divides the series into non-overlapping segments of length *n* and calculates the mean and standard deviation of each segment for increasing segment lengths *n*. The relation between the standard deviation *SD* and *n* is given by

On a bi-logarithmic plot of *SD* versus *n*, the slope of the linear regression provides the scaling exponent *H-1*, where *H* is the Hurst coefficient.

Scaled Windowed Variance Method (SWV)

Scaled Windowed Variance analysis (Cannon et al, 1997) divides the series into non-overlapping segments of length *n* and calculates the average standard deviation for all segments. This is repeated for increasing values of *n*. A bi-logarithmic plot of the average SD versus *n* provides an estimate of the scaling exponent. A refined method, Signal Summation Conversion (Eke et al, 2000) applies the scaled windowed variance to the cumulative sum of the series.

Fractional ARIMA analysis (fARIMA, or ARFIMA)

Fractional ARIMA analysis is an extension of ARIMA analysis (Box & Jenkins, 1970). Wagenmakers et al (2004) propose a competitive testing approach, which compares a range of ARIMA models to a range of fractional ARIMA (ARFIMA) models using the Akaike Information Criterion (AIC, Akaike, 1974) to determine the best model for a given data series. ARFIMA(*p,d,q*) models differ from ARIMA(*p,d,q*) models in that ARIMA(*p,d,q*) models only allow integer values of the differencing parameter *d*, whereas ARFIMA(*p,d,q*) models allow fractional values of the integration parameter *d*. In simple terms, if ARIMA(0,0,0) models white noise with spectral slope $\alpha=0$, and ARIMA(0,1,0) models a random walk with spectral slope $\alpha=-2$, it follows that

1/*f* noise, with a spectral slope of $\alpha \approx -1$, has an integration parameter $0 < d < 1$ (Wagenmakers et al, 2004).

Spectral Classifier (SC)

The Spectral classifier method (Thornton & Gilden, 2004) compares the power spectrum of the experimental series, obtained using power spectral density analysis (PSD) , against power spectra of a long-range correlation series (specifically, fractional Brownian motion with added Gaussian noise), and a short-range correlation series (an autoregressive moving-average ARMA model) using log-likelihood. This approach is similar to the fractional ARIMA approach proposed by Wagenmakers et al (2004), with experimental series tested against alternative series with known long-range and short-range correlation; the distinction is that the spectral classifier method utilises the power spectral density analysis (PSD) method, whereas the fractional ARIMA method utilises linear time series analysis methods.

Appendix B

Feedback provided by participants in Musha, Katsurai and Teramachi (1985)

Participant 1

(experience with piano for 12 years, drum for 2 years)

"I was relaxed during [synchronised tapping] because I had the ticking standard to follow, but was very tired after the [free tapping]."

Participant 2

(dancing experience for 4 years, and teaches dancing)

"During the [free tapping], I was entirely relaxed and did not feel tired. During the [synchronised tapping] I became sleepy in the 4^{th} phase and very tired in the test."

Participant 3

(experience in Chinese dramas and singing, and enjoys ping pong; reports lack of endurance)

"During the [free tapping], I felt tired in the first and last (7^{th}) phase of tapping, enjoyed thinking of friends, the motherland (China), musical melodies, etc., during the tapping, but felt tired in the breaks. In the [synchronised tapping], I became nervous and tired in my keeping pace with the metronome, and was unable to concentrate myself in thinking of a certain subject."

Participant 4

(experience playing guitars for 8 years)

"During the [free tapping], I felt very tired in the 1st phase, but it was easy after that. As I was looking at twinkling red lamps on the computer panel which were synchronous to the metronome, I think my tapping was 4-beat. In the [synchronised tapping], I felt easy because I had the pace maker. I was thinking nothing."

Participant 5

(no training or experience with music)

"During the [free tapping], my tapping rhythm was getting faster, which I was unable to control, and I felt as if my brain had been empty. After the [synchronised tapping], I felt more tired than after experiment B."