Some implications of chaos theory for the genetic analysis of human development and variation

Lindon J Eaves [1], Katherine M Kirk[2], Nicholas G Martin[2] and Robert J Russell[3]

[1]*Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, Virginia, USA*
[2]*Queensland Institute of Medical Research, Brisbane, Queensland, Australia*
[3]*Center for Theology and the Natural Sciences, Berkeley, California, USA*

Non-linear epigenetic processes are a potential underlying source of phenotypic differences in development. Simulation studies of twin pairs using simple non-linear development models characterised by chaotic or near-chaotic behavior are presented. The effect of chaotic processes on correlations is to lower them from their initial values, but high initial correlations are affected much less by chaotic and near-chaotic processes than intermediate correlations. Therefore, we would predict that traits affected by chaotic processes would have high MZ and low DZ twin correlations and this is reminiscent of certain traits such as EEG spectra. However the much more frequent observation of MZ correlations close to twice their DZ counterparts would suggest that the role of chaos in development is quite limited.

Keywords: chaos, non-linear, processes, twins, epigenesis

Introduction

Molenaar *et al*[1] have drawn attention to the possibility that chaos theory might provide a framework for describing the epigenetic processes which generate the 'environmental' differences within pairs of identical twins (E1), and the variation between duplicate structures in the same individual (eg between numbers of sternopleural chaetae on different sides of the same fly, or melanocytic naevi (moles) on left and right sides of the human body). This important insight suggests that the inherent implications of certain non-linear processes may offer an entirely 'deterministic' account in terms of endogenous developmental processes of findings that psychologists and others might seek to explain by reference to external environmental factors.

This paper presents the results of simulation studies which allow us to:

1) develop some implications of chaos theory for the analysis of human behavioral development when this is conducted within the framework of current linear structural models;

2) point out some further consequences for genetic modeling that were not discussed by Molenaar *et al*;

3) evaluate how far chaos theory might be applicable to the types of trait currently studied by behavioral geneticists.

Logistic equation – a simple developmental model capable of producing chaos

Wildman and Russell[2] consider the behavior of the discrete logistic equation

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

on the unit interval [0,1]. The equation represents x_{n+1} (the $n + 1$ th state of variable x) as a non-linear function of the value of x in the previous (nth) state, x_n. Hence, the value of x at any time n can be iteratively determined from the starting value, x_0, and the sequence of values taken by x is known as its *orbit*. The behavior of the logistic equation varies greatly, depending on the value of the constant k, but the iterated sequence of x values remains within the unit interval for $k \leq 4$. For values of $k < 3$, the limit value or equilibrium state of x is a constant, irrespective of the starting value x_0. For $k \leq 1$ the limiting value of x is always zero, whilst for $1 < k \leq 3$ the process converges to a non-zero value which depends only on k. A value towards which x converges is known as an *attracting point*, and if it remains unchanged by iteration (ie $x_{n+1} = x_n$ for a given k after a certain number of iterations) it is known as a *fixed attracting point*. All attracting points are fixed attracting points for $k < 3$.

As k becomes greater than 3, the logistic equation enters a region of *period doubling*, and no longer

converges to a single limit. Instead, the sequence of values for x splits or *bifurcates*, oscillating between convergence towards more than one distinct attracting point. At first there are only two distinct limits but as k increases, the number of attracting points continues to double. These attracting points are known as *periodic points*, as they are transformed into themselves by the logistic equation after an exact number of iterations known as a *limit cycle*. For the first time, the initial starting value x_0 has some impact on the value of x at time n, determining the order in which x approaches the various periodic attracting points. It is important to note that this region of periodic doubling is not chaotic, but forms a transitional boundary between linear and chaotic behavior.[3]

True chaotic behavior of the logistic equation begins at a value of k approx $3.5699 = k_{crit}$, the *accumulation point*. In this region, characterised by a mixture of stable periodic and chaotic behavior, both the starting condition x_0 and the value of k are important in determining the value of x at time n. As discussed by Wildman and Russell,[2] for a fixed value of k, even slight variations in initial conditions are sufficient to cause extremely different iterated sequences. For $k = k_{crit}$, the orbit of x is confined to narrow bands located between 'constraints'. As k increases above k_{crit}, these bands widen until orbits for $k = 4$ cover the entire unit interval. Beyond, the chaotic behavior of the logistic equation is more complex, and for many values of x_0 the result of iteration is a value outside the unit interval $[0,1]$.

Modeling differences between twins

The parameter k, and the form of the logistic equation, may be considered as defining the 'epigenetic rules' which govern the trajectory of the phenotype x over time. Thus, they are the rules governing the development of trait x. Since the initial state of the logistic equation has no bearing on the final limit of convergence for $k \leq 3$, a value of k in this range would result in individuals' phenotypes converging to the same value over time. A value of k greater than 4 in many cases results in the orbit of x departing from the unit interval of $[0,1]$ and accelerating towards infinity.[2] Since the range of possible phenotypes occurring in humans does not extend to infinite values, interest in the behavior of the logistic equation from the standpoint of the study of individual differences is only for $3 < k \leq 4$.

Modeling of the correlation between members of a twin pair via the logistic equation has been undertaken using FORTRAN, with the inclusion of NAG subroutines G05CBF and G05DDF for generation of pseudo-random real numbers with a normal distribution.[4] The model specifies the phenotypic mean μ, the within-pair variance σ_w^2, the between-pair variance σ_b^2, constant k ($3 < k \leq 4$), the number of twin pairs over which results are averaged (500), and the number of iterations (50). The total variance ($\sigma_w^2 + \sigma_b^2$) is standardised to 1. Phenotypic 'starting values' for each twin are generated using the NAG subroutines with μ, σ_w^2 and σ_b^2 as input parameters. The logistic equation is then iterated to obtain the final phenotypic values at time n. Since results are simulated for a large number of twin pairs, the overall correlation between twins at time n for those starting conditions can then be estimated.

Results

Figure 1 illustrates the phenotypic correlation between members of twin pairs for $k = 3.4$, a value within the bifurcation regime $3 < k < k_{crit}$. For each combination of within-pair and between-pair variance, the correlation at time n of the simulated twin pairs has been plotted, with relatively smooth convergence to the final correlation value in each case. For very small within-pair variance (0.01), the correlation between twins remains high (approx 0.85), whereas for even a within-pair variance of 0.2 the correlation between twins is less than 0.4.

Behavior of the logistic equation for values of k just above the accumulation point is demonstrated in Figure 2, for $k = 3.57$. As in Figure 1, a very high initial correlation between twins is maintained as iteration proceeds, but even a slightly lower initial correlation of about 0.8 is rapidly degraded to less than half its original value. However, with a limited number of values that the logistic equation may take for a given set of initial conditions due to banding, some correlation between the twins' phenotypes is preserved despite the chaotic nature of the logistic equation. The principal difference between Figures 1 and 2 does not occur in the values towards which the twin correlations converge, but in the manner in which they do so. In this case, the correlation values obtained for given within-pair and between-pair variances oscillate around the limiting value, rather than approaching it smoothly.

The final figure, Figure 3, demonstrates the behavior of the logistic equation as k increases to a value of 3.84. In this case, the 'observed' twin correlations are much lower than the corresponding values for $k = 3.57$ and the same initial conditions. This is due to the fact that the logistic equation values are not constrained to such narrow bands. At these higher values of k, iteration of the logistic equation causes the phenotypic correlation between twins to decrease rapidly, regardless of its initial value.
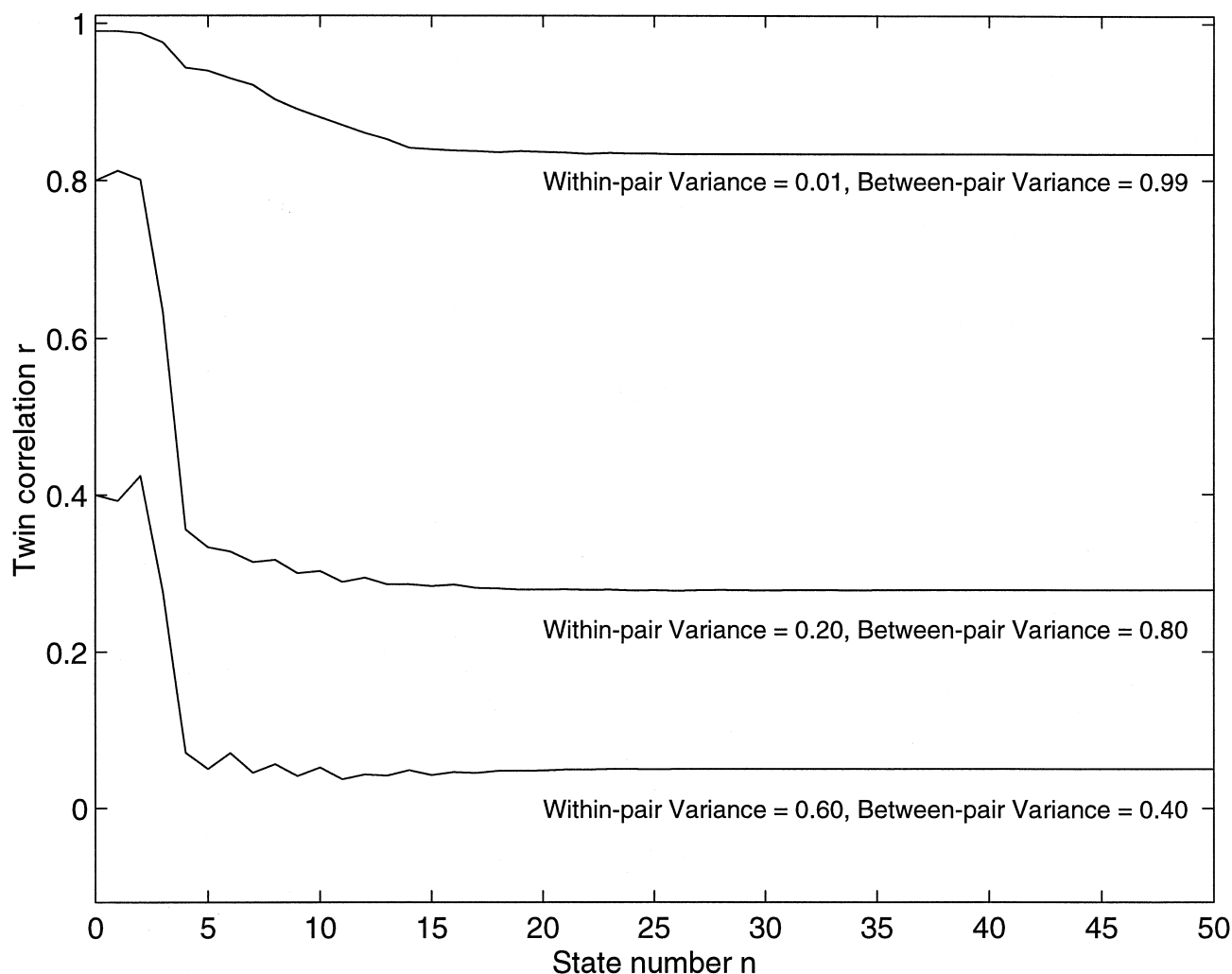
Figure 1 Logistic equation modeling of twin correlations, k = 3.40

Discussion

We note that Molenaar *et al*[1] attribute to the epigenetic consequences of 'chaos', a status which is additional to, and independent of, the traditional two-fold categorisation of the primary causes of variation into 'genetic' and 'environmental' factors. This distinction is likely to cause confusion. 'Chaos' is not an alternative to 'environment' but, as the authors imply elsewhere, a source of environmental variation which more naive investigators might seek to 'explain' in terms of other more conventional environmental factors such as differences in the treatment of MZ twins. What the 'chaos model' does, however, is alert behavioral geneticists to the fact that the search for specific 'environmental' variables might be unrequited.

Our simulations, however, allow us to be more specific in a number of directions. Firstly, when we characterise development by a simple non-linear model which is behaving chaotically or near-chaotically, we find that *small variations in initial conditions* (eg small quantitative differences between MZ twins at the molecular level such as the degree of methylation of a particular gene in a particular tissue) *will have consequences at the phenotypic level which look like occasion-specific environmental effects*. Put another way, what linear models identify as short term environmental influences ('unreliability' etc) and others may seek to explain in terms of changes in the actual environment over time, may be nothing more than the inherent and virtually unpredictable specific behavioral fluctuations of uniquely different individuals under a non-linear epigenetic process. Thus, early environmental
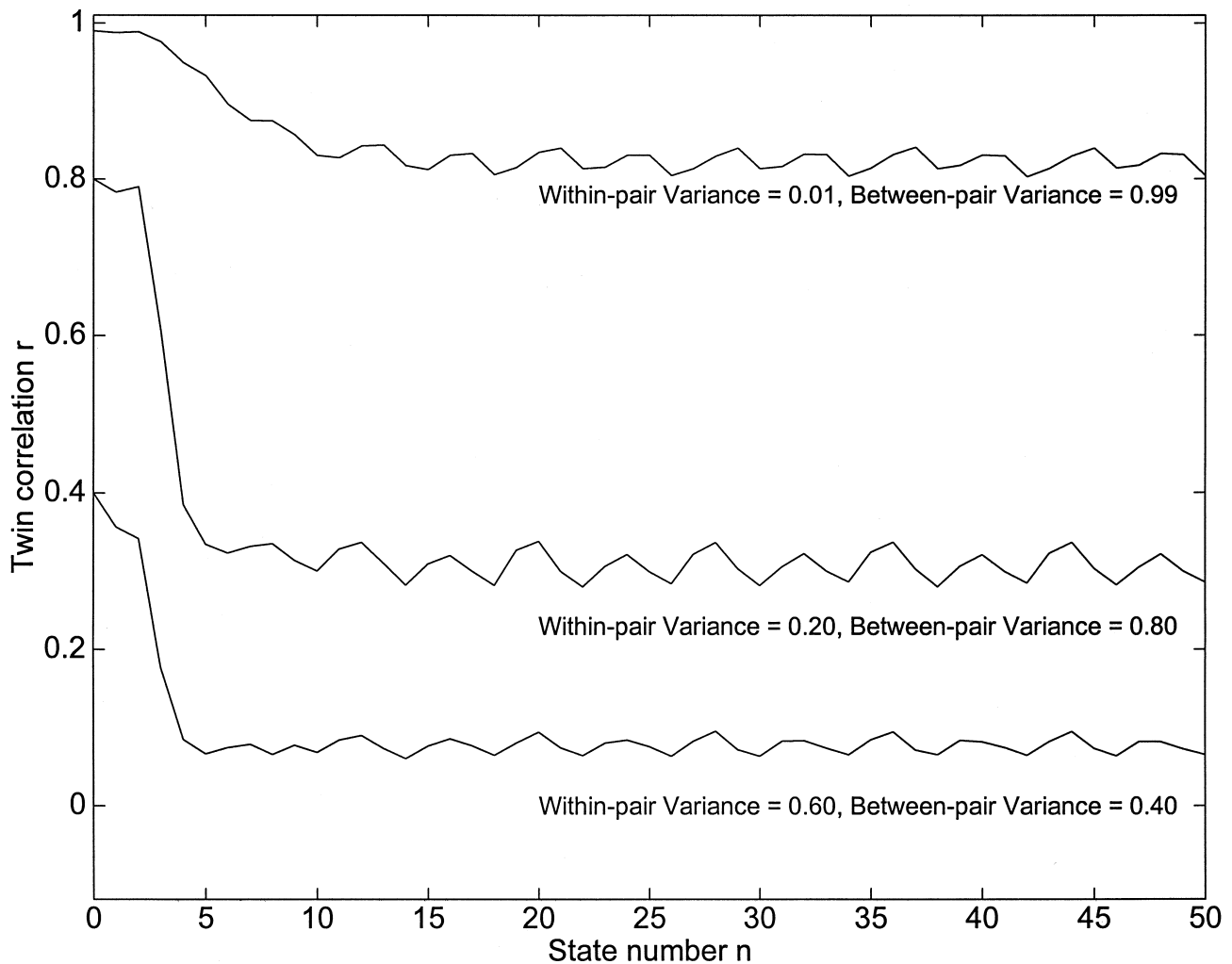
Figure 2 Logistic equation modeling of twin correlations, k = 3.57

differences may, under a 'chaos' model, be sufficient to explain a series of very short-term environmental differences which others might be tempted to explain by specific correlated short-term changes in exogenous variables. The 'chaotic' model thus leads to an alternative research paradigm to that presupposed by the attempt to explain short-term behavioral fluctuations in terms of external factors.

Secondly, the epigenetic process implied by our simple non-linear model can also capitalise on *genetic* differences in the initial conditions to give rise to what, under a linear genetic model, might be treated as 'occasion specific' genetic effects. Uncritical application of the linear model might tempt investigators to search for specific loci which were 'switched on' in a particular developmental 'window' when, in fact, there are no particular loci responsible for a particular age-specific genetic effect.

Finally, the precise consequences of a chaotic or near-chaotic process for the correlation between relatives (eg MZ and DZ twins) at any specific point in development will depend on the correlation between relatives for the initial conditions. Generally, twin correlations less than about 0.5 in the initial conditions very quickly produce correlations around zero between twins in subsequent cross-sectional measures, as observed in the simulations presented here. On the other hand, high initial correlations (upwards of 0.9) can yield superficially stable (though minutely fluctuating) non-zero correlations between twins on later measures (see Figure 1 and 2). One important indication that a given system is nonlinear, therefore, would be the demonstration that DZ correlations are close to zero when the MZ correlations are still far too high. Explanations in terms of dominance, epistasis[5] or 'emergenesis',[6] do not get right to the heart of the matter because they
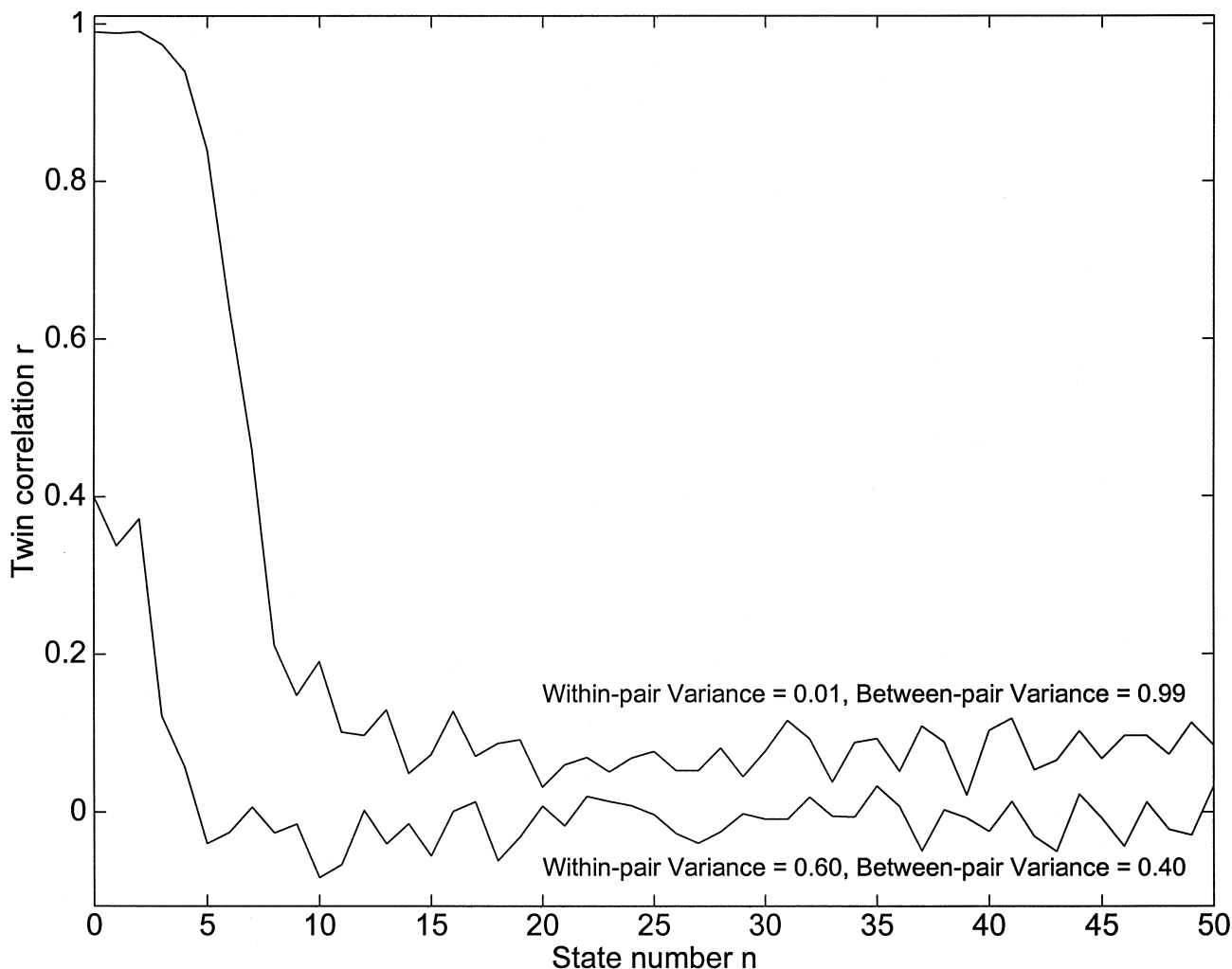
Figure 3 Logistic equation modeling of twin correlations, k = 3.84

are still modeling the process in terms of interactions between individual components of a complex system. The 'chaos' model may generate exactly the kinds of correlations sometimes reported (very low DZ correlations) without recourse to complex parameterisations even when variation in the initial correlations are almost entirely explained by additive genetic factors.

How far models of this kind will turn out to be important in behavior genetics will depend on a number of factors. There are so few well-explored models of human behavioral development that it is impossible to determine general principles of development, if any. It is apparently the case that a large fraction of within family environmental variation is relatively short-lived. This would favor an explanation in chaotic terms. However, the correlations for behavioral outcomes tend not to be grouped generally into very high MZ correlations (greater than 0.8, say) and very small DZ correlations (say less

than 0.1). Although there are examples of such in the literature (see, for example, some published twin correlations for EEG measures), we find the whole range of MZ and DZ correlations, with many MZ correlations for personality and attitudes stabilising in the 0.4–0.6 range rather than higher, and many DZ correlations for cognitive variables finishing up in the same range rather than lower. Thus, although 'chaos' may be an important element in development, its effects may not be universal, otherwise the developmental data would appear differently. It may be that some parameters of the EEG will be the best place to start looking for chaotic effects, as Molenaar *et al*[1] imply, but if our simulations offer a legitimate guide to what they might find, we should not expect, as they also suggest, to find the 'chaotic' element confined purely to estimates of the within family environmental variation. Rather, we may anticipate non-additive genetic effects of a magnitude that cannot easily be encompassed by reasonable values

of dominance and epistatic components of variance under the 'linear' genetic model.

We cannot say how far, or under what circumstances, natural selection would 'abhor chaos' and under what conditions natural selection would 'favor' a chaotic epigenetic system over alternatives. Molenaar *et al* rightly direct our attention to the literature on the genetic control of sensitivity to the environment as the best source to begin answering this important question. Citing Mather,[7] they conclude that dominant genetic control of stability/instability appears to be in the direction of increasing stability (heterosis being in the direction of stability and inbreeding depression tending to lead to instability). If such a finding is widespread, we should probably conclude that chaos is the exception rather than the rule as a viable evolutionary strategy.

References

1 Molenaar PCM, Boomsma DI, Dolan CV. A third source of developmental differences. *Behav Genet* 1993; **23**: 519–524.
2 Wildman WJ, Russell RJ. Chaos: a mathematical introduction with philosophical reflections. In: Russell RJ, Murphy N, Peacocke AR (eds). *Chaos and Complexity: Scientific Perspectives on Divine Action*. University of Notre Dame Press: Notre Dame, 1996, pp 49–89.
3 Weisstein EW. *The CRC Concise Encyclopedia of Mathematics*. CRC Press: Florida, 1998.
4 Numerical Algorithms Group. *NAG Fortran Library Manual, Mark 17*. Numerical Algorithms Group Ltd: Oxford, 1995.
5 Eaves LJ. Dominance alone is not enough. *Behav Genet* 1988; **18**: 27–33.
6 Lykken DT, McGue M, Tellegen A, Bouchard TJ Jr. Emergenesis. Genetic traits that may not run in families. *Am Psychol* 1992; **47**: 1565–1577.
7 Mather K. *Genetic Structure of Populations*. Chapman and Hall: London, 1973.