



Additional aspects of third source variation for the genetic analysis of human development and behaviour: a commentary on Eaves *et al*

Peter CM Molenaar and Maartje EJ Raijmakers

University of Amsterdam

Introduction

Eaves *et al*¹ present a series of simulation studies showing some effects of chaotic processes on MZ and DZ twin correlations. More specifically, they use the logistic equation, $x(t + 1) = kx(t)[1 - x(t)]$, where k is a parameter and t denotes discrete time, as a simple developmental model and simulate MZ and DZ twin phenotypic time series under various assumptions concerning the starting value $x(0)$ and using a number of distinct values for the parameter k . In discussing the results thus obtained, Eaves *et al*¹ draw some conclusions with respect to, for example, the possible forms and prevalence of chaotic variation in developmental processes as proposed by Molenaar *et al*.² In the following we present the results of alternative simulation studies that show additional aspects of chaotic variation in the genetic analysis of human development. Finally, we discuss some of the implications of our own results with respect to the interpretations given by Eaves *et al*¹ of their simulation results.

Alternative simulation designs

In each replication of the series of simulation studies by Eaves *et al*¹ it is assumed that the logistic equation describing development is invariant across subjects. That is, in each replication the value of the parameter k is identical across subjects, whilst the starting value varies between subjects. The variation in the starting value $x(0)$ between subjects is described as a linear combination of variance within twin pairs, σ_w , and variance between twin pairs, σ_b . The choice of particular values for σ_w and σ_b defines the within-pair correlation of starting values. The main result in each replication then is a plot of the within-pair correlation as a function of time t . It is found that the value of the within-pair correlation

decreases as a function of time, where the rate of decrease is inversely related to the initial within-pair correlation of $x(0)$.

In view of existing successful simulation studies of embryogenesis by eg Meinhardt,³ it was suggested by Molenaar *et al*² that a plausible alternative design is one in which all subjects obey the same developmental process model, but the between-subjects variation of the parameters in this model is under genetic control. Molenaar⁴ applied this alternative design in the following way. Chaotic phenotypic time series are generated for 250 MZ twin pairs and 250 DZ twin pairs by means of the logistic equation in which the parameter k has a mean equal to 3.5 and heritability equal to 1.0 (between-subjects variance of k is 0.01). In addition the starting value $x(0)$ is 0.1 for one member and 0.2 for the other member of each pair. Hence in this particular simulation design the parameter k in the logistic equation varies between subjects, where this variation is entirely under genetic control, whilst the starting values are fixed between pairs (σ_b is zero). It is then found that the heritability of the phenotypic time series (as assessed by standard maximum likelihood estimation; the likelihood ratio test of the model including an additive genetic factor and a specific environmental factor has $P = 0.99$) is 0.98 at time $t = 5$. In contrast, the estimated heritability decreases to 0.07 at time $t = 1000$ (likelihood ratio test has $P = 0.90$).

After reading the manuscript of Eaves *et al*¹ we carried out the following simulation experiment. Chaotic phenotypic time series are generated for 100 MZ twin pairs and 100 DZ twin pairs by means of the logistic equation in which the parameter k has a mean equal to 3.8 and heritability equal to 1.0 (between-subjects variance of k is 0.01). In addition the starting value has a mean of 0.5 and heritability equal to 0.999999 (between-subjects variance of $x(0)$ is 0.01). Hence in this experiment both the parameter k and the starting value varies between subjects, where this variation is completely under genetic control for parameter k and almost completely under genetic control for the starting value. In agreement with the general pattern of results obtained by Eaves *et al*,¹ it is found that the phenotypic MZ within-pair correlation is high at an early stage, whilst the

Correspondence: Peter C. M. Molenaar, Faculty of Psychology, University of Amsterdam, Roetersstraat 15, room 1010, 1018 WB, Amsterdam, The Netherlands; E-mail: op_molenaar@macmail.psy.WVA.NL

Received 4 December 1998; accepted 11 January 1999

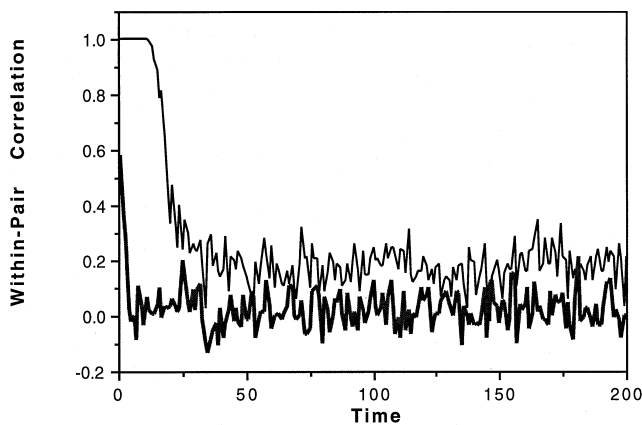


Figure 1 Within-pair correlation of MZ twins (upper trace) and DZ twins (lower trace) as a function of time

phenotypic DZ within-pair correlation is about zero. Figure 1 shows the phenotypical MZ and DZ within-pair correlations thus obtained for $t = 1, \dots, 200$. The model including an additive genetic factor and a specific environmental factor does not fit (likelihood-ratio test has $P = 0.00$). However, at a later stage the phenotypic MZ within-pair correlation also has decreased significantly. At $t = 50$ the estimated heritability is 0.12 (likelihood-ratio has $P = 0.78$).

A new estimation scheme

One intriguing aspect of the results presented in the previous section is that after a finite interval of time the estimated heritability of phenotypic chaotic time series approaches zero. Yet the phenotypic series have been generated by means of a developmental model in which the parameter k has heritability equal to 1.0 and the initial value $x(0)$ has heritability of 0.999999. In some respects the strong genetic structure of the developmental model becomes 'invisible' in a straightforward quantitative genetic analysis of its behaviour. This special feature of chaotic developmental processes has been discussed at some length by Burgess and Molenaar.⁵ Below we suggest a new estimation scheme by which the genetic structure in developmental processes can be recovered.

Consider the same simulation experiment as described at the end of the previous section. That is, chaotic phenotypic time series of length 200 are generated for 100 MZ twin pairs and 100 DZ twin pairs by means of the logistic equation in which the parameter k has a mean equal to 3.8 and heritability equal to 1.0 (between-subjects variance of k is 0.01). In addition the starting value has a mean of 0.5 and heritability equal to 0.999999 (between-subjects variance of $x(0)$ is 0.01). Figure 1 shows the phenotypy-

ical MZ and DZ within-pair correlations thus obtained. We now take the final half of each generated phenotypic time series, where the estimated heritability at each time point $t = 101, 102, \dots, 200$, is about zero. The basic data set thus obtained is an ensemble of univariate time series, $x_{in}(t)$, $t = 101, 102, \dots, 200$; $i = 1$ (MZ) or $i = 2$ (DZ); $n = 1, 2, \dots, 200$.

Now the parameter k in the logistic equation is estimated for each single-subject time series $x_{in}(t)$ in the ensemble. This can be accomplished by means of special time series analysis techniques (extended Kalman filtering) as explained in Molenaar.⁶ In this way, an estimate of k is obtained for each subject. Next the heritability of k is determined in the usual quantitative genetic analysis described above. It then is found that the estimated heritability of the estimated parameter k in the logistic model is 1.0 (likelihood ratio has $P = 0.19$).

Discussion

Perhaps the most general conclusion which can be drawn from the simulation studies of Eaves *et al*¹ and those described in the previous section is that for the logistic developmental model the within-pair correlation between initial values $x(0)$ at time $t = 0$ decreases as a function of time. In a quantitative genetical analysis of this logistic developmental model, it will be found that the heritability of the phenotypic series also decreases as a function of time, at least at sufficiently long periods of time t . In itself this is a significant result: the genetic structure which is present at the start of the developmental process becomes increasingly 'invisible' in quantitative genetical analyses of the phenotypic time series at later times. From this result an apparent paradox was derived by Burgess and Molenaar⁵ according to the following reasoning. The effects of genetic information on neural ontogenesis are mediated through nonlinear dynamic processes with self-organising properties (the total informational content of the human genome is far too small to control the creation of all synaptic interconnections in the brain). Given the 'disappearance' with time of genetic structure initially present in nonlinear dynamic processes, one would expect to find low heritabilities of (eg cognitive) behaviour dependent on mature brain structures. In contrast one finds substantial heritabilities for various cognitive and other behaviour in empirical quantitative genetic analyses.

As to the details in which the within-pair correlations decrease as a function of time, we notice some interesting differences between the simulation results of Eaves *et al*¹ and ours. Eaves *et al*¹ indicate

that in their simulation experiments a high initial within-pair correlation of $x(0)$ decreases much more slowly as a function of time than a medium-valued initial within-pair correlation. This implies that within some time interval shortly after $t = 0$, DZ within-pair correlation of $x(t)$ can be close to zero, whilst MZ within-pair correlations remain rather high. Because this distinct pattern of MZ and DZ within-pair correlations is not commonly observed in empirical quantitative genetic analyses, Eaves *et al.*¹ suggest that the effects of chaotic processes in real developmental systems may not be universally present. We would like to qualify this suggestion by Eaves *et al.*¹ in the light of the results obtained in the alternative simulation experiments described in the previous section.

First and foremost, it is a distinguishing feature of chaotic processes that differences in the initial condition, however small, will increase as the process evolves in time. This implies that after a sufficient amount of time any initial within-pair correlation will vanish, although at different rates for high initial within-pair correlations (MZ pairs) in comparison with low initial within-pair correlations (DZ pairs). Indeed, this is what we observe in our simulation experiments if t increases to 100 or more. More specifically, for $t > 100$ both MZ and DZ within-pair correlations of $x(t)$ are close to zero, quantitative genetic analyses based on the standard model including additive genetic and specific environmental factors yield acceptable model fits, and estimated heritabilities are close to zero. Consequently we discern the following general scenario for the evolution of chaotic phenotypic time series: during an initial time interval (which may be very short or non-existent) the standard additive genetic model yields acceptable fits to the data and estimated heritabilities can be high. During the next interval of intermediate times the standard additive genetic model yields unacceptable fits to the data and non-additive genetic parameters (dominance or epistasis) must be added. Finally, after a sufficient amount of time the standard additive genetic model yields acceptable fits to the data, but estimated heritabilities are close to zero.

The detailed ways in which a chaotic developmental process evolves in time depends, among other things, upon the type of dynamics (in discrete time for instance the logistic equation, Ikeda equations etc; in continuous time for instance the Lorenz equations, nonlinear reaction-diffusion equations etc), the time that has elapsed since the start of the developmental process (see discussion above), the parameter values in the dynamic equations describing the developmental process, and the genetic structure underlying the phenotypic time series (see previous sections). Hence one should expect a broad

variety in the details of time-dependent changes in within-pair correlations, as function of differences in model variants and simulation designs. In the first alternative simulation design described earlier, there is genetic variation for the model parameter k of the logistic equation, whereas in the second simulation experiment the genetic structure pertains to both the model parameter k and the starting value $x(0)$. The time-dependent decrease of initial within-pair correlations are found to be much more similar for MZ and DZ twin pairs in the first simulation experiment than in the second. It would seem safe to conclude that a more precise delineation of the boundaries of domains of generalisation for results obtained with particular model variants and simulation designs requires much further study.

The introduction of the new estimation scheme in which first each single-subject time series is analysed, yielding individual estimates of the parameter k in the logistic model, and then the individual estimates of k are subjected to a quantitative genetic analysis, shows that the genetic structure initially present at the start of a chaotic developmental process does not really disappear as this process evolves in time (hence the paradox referred to earlier is only apparent). It only becomes 'invisible' in a straightforward quantitative genetic analysis of the phenotypical time series, but not after Kalman filtering. This result is expected to have far-reaching consequences for the application of chaos theory in the context of quantitative genetics. Some of these consequences are further elaborated in current simulation studies, to be reported in a paper by Molenaar and Raijmakers⁷ (in preparation). For the moment it is worth indicating a relationship of the proposed new estimation scheme with the standard approach in quantitative genetic analyses of EEG. In a quantitative genetic analysis of EEG it is also standard to analyse first each single-subject EEG signal and derive individual spectral estimates characterising its time-dependent behaviour, after which the individual spectral estimates are subjected to quantitative genetic modelling, see for instance Beijsterveldt *et al.*⁷ Clearly, the general procedure followed in the quantitative genetic analysis of EEG data resembles our new estimation procedure for chaotic time series.

In closing, we address a few more technical issues related to simulation experiments with chaotic dynamic models. Firstly, the real time interval between consecutive values of t , ie the basic time step $\delta t = (t + 1) - t$, in the logistic model (or any other dynamic model in discrete time) is undefined. More specifically, depending on the natural time scale of the empirical process at hand, δt can be a millisecond, a second, an hour, a month, etc. If, for instance, the logistic equation is used as a model of

biochemical processes going on in the evolving neocortex of a newborn baby, then δt will be of the order of seconds. Accordingly, if the baby becomes about one year old then the upper time index of the logistic equation concerned will be in the range of $t = 10^7$. For such large values of t all within-pair correlations initially present at $t = 0$ will have long ago vanished completely. Hence, depending on the natural time scale of particular developmental processes, the special transient pattern of rather high MZ within-pair correlations in combination with low DZ within-pair correlations may be present in real time only for a short span of time, after which this pattern decays to one in which all phenotypical within-pair correlations are about zero. Secondly, we would point out that computer simulations of chaotic processes are vulnerable to the inevitable finite precision of real number representations on the hardware. One implication of this is the following. Suppose the logistic equation is assigned a particular starting value $x(0) = x_0$. The value of x_0 is represented in the computer with finite precision, for instance by specifying this number up to 7 decimals: $x_0 = 0.1234567$. This implies that x_0 thus specified with (of necessity) finite precision is in reality the real value number $x_0 = 0.1234567000000000\dots$, where the ellipses denote an infinite sequence of zeros. Now consider the possibility of realising exactly this starting value $x_0 = 0.1234567000000000\dots$ in a real developmental process, for instance the biochemical process alluded to above in which $x(t)$ denotes the (relative) concentration of an enzyme. To ensure that the starting concentration of this enzyme is exactly $x_0 = 0.1234567000000000\dots$ would require an infinite amount of information (energy) and hence is biologically implausible. The consequence of this is that 'equal' starting values in real biological processes will always have to differ by infinitesimal amounts, depending on the amount of information and effort which evolution has spent in fixing such starting values. Hence, whilst it is an easy exercise to define equal starting values in finite precision representations during a computer simulation of dynamic models, it will be impossible to do so for real value phenotypes in real biological systems. Then, if the real biological system obeys a

chaotic dynamic regime, these inevitable infinitesimal differences in starting values will eventually increase to such magnitudes that all within-pair correlations initially present will disappear as time proceeds.

We end this commentary by acknowledging the importance of the simulation studies carried out by Eaves *et al.*¹ In our view these simulations provide a clear window on the intricacies created by the workings of chaotic developmental processes. Note that, in contrast to the series of simulations carried out by Eaves *et al.*,¹ each of our illustrative simulation experiments only involves a single replication and hence should be repeated using a full Monte Carlo design with a sufficient number of replications in each condition. Given this qualification, both our preliminary results and the results obtained by Eaves *et al.*¹ should warn quantitative geneticists that straightforward application of their methods to phenotypical data may lead to conclusions about genetic and/or environmental effects which are not warranted where the developmental processes under scrutiny are (or have been) chaotic. We hope and expect that more simulation studies along these lines will follow.

References

- 1 Eaves LJ, Russell RJ, Kirk KM, Martin NG. Some implications of chaos theory for the genetic analysis of human development and variation. *Twin Res* 1999; **2**: 43–48.
- 2 Molenaar PCM, Boomsma DI, Dolan CV. A third source of developmental differences. *Behav Genet*; **23**: 519–524.
- 3 Meinhardt H. *Models of Biological Pattern Formation*. Academic Press: London, 1982.
- 4 Molenaar PCM. Introduction to chaos theory and bifurcation theory. Paper presented at the Workshop in Genetics of Non-linear Dynamic Systems, University Park, PA, USA, 27 May–1 June, 1998.
- 5 Burgess RL, Molenaar PCM. Commentary. *Hum Devel* 1995; **38**: 159–164.
- 6 Molenaar PCM. *Dynamic Latent Variable Models in Developmental Psychology*. In: von Eye A, Clogg CC (eds). *Analysis of latent variables in developmental research*. Sage: Newbury Park, CA, 1994, pp 155–180.
- 7 Beijsterveldt CEM van, Molenaar PCM, Geus EJC de, Boomsma DI. Genetic and environmental influences on EEG coherence. *Behav Genet* 1998; **28**: 443–453.