MEASUREMENT, 2(4), 201–218

FOCUS ARTICLE

A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever

Peter C. M. Molenaar
Department of Psychology
University of Amsterdam

Psychology is focused on variation between cases (interindividual variation). Results thus obtained are considered to be generalizable to the understanding and explanation of variation within single cases (intraindividual variation). It is indicated, however, that the direct consequences of the classical ergodic theorems for psychology and psychometrics invalidate this conjectured generalizability: only under very strict conditions—which are hardly obtained in real psychological processes–can a generalization be made from a structure of interindividual variation to the analogous structure of intraindividual variation. Illustrations of the lack of this generalizability are given in the contexts of psychometrics, developmental psychology, and personality theory.

Keywords: idiography, interindividual variation, intraindividual variation, ergodicity, psychological processes

When at the start of the new millennium scientists of all kinds of disciplines were asked what they considered to be the single most important scientific breakthrough of the 20th century, the majority chose Brownian motion. Indeed, the construction of a model for Brownian motion by Einstein in 1905 led to a revolution in science, not

Requests for reprints should be sent to Peter C. M. Molenaar, Dept. of Psychology, University of Amsterdam, Roetersstraat 15, Room 504, 1018 WB Amsterdam, The Netherlands. E-mail: p.c.m.molenaar@uva.nl

only in physics (statistical mechanics, quantum physics), but also in biology (pattern formation, evolutionary processes), mathematics (stochastic calculus), economics and, of course, psychology. Modern psychology is saturated with probability models and statistical techniques. Yet, surprisingly, psychologists have mainly been interested in part of the results of the stochastic revolution initiated by Einstein. Although the model of Brownian motion explicitly pertains to the random (stochastic) time-dependent behavior of a single particle (system), it appears that such a pure N = 1 time series perspective is almost completely lacking in psychology. Attention in psychological research is almost exclusively restricted to variation between individuals (interindividual variation [IEV]), to the neglect of time-dependent variation within a single participant's time series (intraindividual variation [IAV]).

As will be explained later, psychology as an idiographic science restores the balance by focusing on the neglected time-dependent variation within a single individual (IAV). It brings back into scientific psychology the dedicated study of the individual, prior to pooling across other individuals. Each person is initially conceived of as a possibly unique system of interacting dynamic processes, the unfolding of which gives rise to an individual life trajectory in a high-dimensional psychological space. Thus, bringing back the person into scientific psychology, it can be proven that his or her return is definitive this time. Classical theorems in ergodic theory, a branch of mathematical statistics and probability theory, show that most psychological processes will have to be considered to be nonergodic. For nonergodic processes, an analysis of the structure of IEV will yield results that differ from results obtained in an analogous analysis of IAV. Hence, for the class of nonergodic processes (which include all developmental processes, learning processes, adaptive processes, and many more), explicit analyses of IAV for their own sakes are required to obtain valid results concerning individual development, learning performance, and so forth.

The foundational issue at stake concerns the relation between the structure of IEV and the structure of IAV. Precise definitions of the terms of this relation will be given later, but first a heuristic illustration is given. Suppose that a standard factor analysis of p-variate measurements obtained with a sample of N participants yields a solution involving q latent factors, where each factor can be assigned a definite interpretation. This factor solution then constitutes a description of the structure of IEV, because the covariance matrix from which it is obtained is computed in a particular way, namely by taking the outer product[1] of the p-variate vector of scores of each participant and then averaging these outer products across the N participants.

[1]Let $y_i = [y_{i1}, y_{i2}]'$ be a two-variate vector of scores for participant i, where the prime symbol (′) denotes vector transposition. Hence, y_i denotes a two-variate column vector. Then, the outer product of y_i is the product of y_i with its transpose, the row vector y_i'. This yields a (2, 2)-dimensional matrix, the (j, k)-th element is the product $y_{ij} * y_{ik}$, j, k = 1, 2. Hence, this matrix has the squares of the individual scores on the diagonal and the cross-products of the individual scores as off-diagonal elements.

If N participants would yield the same vector of scores, then the covariance matrix would be the zero matrix and no factor solution could be obtained. Hence, the manner in which the N participants yield distinct score vectors (i.e., the structure of IEV of these score vectors) provides the information on which the factor analysis is based. Next, suppose that, under the same conditions as before, one of these N participants, S, is repeatedly measured on T consecutive occasions, thus yielding a p-variate time series of score vectors. Take the outer product of the p-variate score vector of S at each occasion and then average these outer products across the T occasions. The covariance matrix thus obtained describes the variation of score vectors within a single participant, S, and can again be subjected to factor analysis (Molenaar, 1985). If S would yield the same score vector at all T occasions, then this covariance matrix would be the zero matrix and factor analysis would be impossible. Hence, the manner in which the score vectors of S vary across the T occasions (i.e., the structure of IAV of these score vectors), provides the information on which factor analysis of the covariance matrix of S's time series is based.

For this particular illustrative situation, the foundational issue involves the following question: Under which conditions can we expect that the factor solution of the covariance matrix characterizing the IEV across N participants is identical to the factor solution of the covariance matrix characterizing the IAV of S across T occasions? More specifically, under which conditions will factor analysis of the p-variate time series of S again yield q factors having the same definite interpretation as in the factor solution for the p-variate score vectors of the N participants?

One might be of the opinion that there should exist some lawful relation between the structures of IEV and IAV, at least in most circumstances and for the majority of psychological processes. For instance, one might point out that the single participant S introduced earlier (the IAV of which is factor analyzed) also belongs to the sample of N participants, the IEV of which obeys a standard q-factor solution. I will show, however, that this attractive opinion is not correct. More specifically, I employ mathematical–statistical theorems obtained in ergodic theory to clearly state the conditions that are necessary (but sometimes not sufficient) for a process to be ergodic. Only for ergodic processes is there a relation (more specifically, an equivalence at asymptote) between the structures of IEV and IAV. The necessary conditions for ergodicity, however, are rather strict and will not be obeyed by many psychological processes. For the latter nonergodic processes, no asymptotic equivalence relation between the structures of IEV and IAV exists (hence, letting the number N of participants approach infinity will be to no avail to correctly identify the analogous structure of IAV and, vice versa, letting the number of time points T approach infinity will not help to correctly identify the structure of IEV). I will explain a few of the far-reaching consequences of nonergodicity at both theoretical and applied levels. In particular, I argue that test theory, yielding the formal and technical underpinning of psychological test construction (Lord & Novick, 1968), gives rise to serious questions regarding its applicability to individual assessment.

Although nonergodic processes exemplify a mild form of heterogeneity (to be specified later), it can be conjectured that real psychological processes are heterogeneous in far more extreme ways. Perhaps each individual person should be considered to be unique in several important ways (one of which is his or her genotype). I refer to some strong theoretical reasons, mainly drawn from mathematical biology, why this conjecture should be taken seriously. To understand how possibly extensive forms of heterogeneity in human populations can go undetected in ongoing psychological research, I show that standard statistical analysis of IEV can be blind for the presence of even extreme forms of heterogeneity.

This article is organized as follows. In the next section some introductory remarks about stochastic processes, a key notion in idiographic psychology, are given. Then follows a section devoted to the definition of ergodicity, being the general property that divides stochastic processes into two distinct classes: those for which the structures of IAV and IEV are equivalent (ergodic processes) and those for which the structures of IAV and IEV are different (nonergodic processes). Next are a number of sections in which the implications of nonergodicity are elaborated within the context of psychometrics and developmental psychology. The closing part of this article summarizes an empirical illustration of the detection of substantial heterogeneity in multivariate psychological time series obtained in a replicated time series design. Using advanced statistical signal analysis techniques, I show that none of the factor solutions, describing the IAV structure of the replicated single-participant time series of personality test score vectors, conforms to the normative (IEV) factor structure of this test.

A final remark on style of presentation. This article is written for an audience that has at least a working knowledge of psychometrics and structural equation modeling. Technical terms and concepts are only explained concisely, sometimes in footnotes. Being a manifesto about the epistemological necessity of idiography, the article does not aim to serve other functions such as covering the existing literature on single-case research. In fact, I know of only one serious research program dedicated to idiographic methodology, namely John Nesselroade's work on P-technique (i.e., standard factor analysis of single-participant multivariate time series). The published literature on idiography occupies only a vanishingly small proportion of our scientific journals, which is an indefensible and unjustified neglect of the facts. The facts are that one should expect the proportion of idiographic research in psychology to be at least 50%. The justification of this expectation is based on solid formal proofs of mathematical–statistical theorems (classical ergodic theorems). To convey this point to the audience I do not need a subtle argument, but a manifesto.

PRELIMINARY REMARKS

A standard dictionary definition of variation is: "The degree to which something differs, for example, from a former state or value, from others of the same type, or from a standard." The degree to which something differs implies a comparison, either between distinct replicates of the thing concerned (IEV) or else between temporal states of the same individual thing (IAV). To simplify matters, we always understand variation to be quantified in terms of variances and covariances, although the gist of our remarks also applies to more general interpretations of the dictionary definition. Hence, what follows the structure of variation, whether IEV or IAV, concerns the second-order moment structure (covariance matrix) of the observations.

The definition of IEV and IAV requires some preliminary discussion of concepts like behavior space and random process. Consider a set of measurable variables yielding a complete description of some domain of interest. Each variable can be represented as a dimension in the multidimensional space spanned by the complete set of variables. We might call this space the phase space. The values of all the variables for an individual participant at a particular point t in time define a point in the phase space. Adding time as an additional dimension to the phase space yields the behavior space. The values of all variables for the same individual participant, realized at consecutive time points, define a trajectory (life history) in the behavior space. According to de Groot (1954), the behavior space contains all the scientifically relevant information about a person. The complete set of life histories of a population of human participants can be represented as a collection of trajectories in the same behavior space.

To define the concept of random process consider, for a given (fixed) single participant, the trajectory in behavior space up to some time t (where the value of t is arbitrary). At t the trajectory characterizing the participant contains all the relevant information about him or her that is available up to, and including, time t. Given this information at t, the prediction of the location of his or her trajectory at the next time point t + 1 (time is taken to proceed in discrete steps for ease of presentation only) will in general not be exact. Even if the time point at which prediction is attempted approaches infinity (and hence the available historical information about the participant's trajectory can increase without bound), exact prediction of the value of this trajectory at the next time point still may not become exact. If this is indeed the case (and it is almost always the case in psychological research), then we consider the trajectory to be the result (realization) of a random process. Hence a random process is characterized by irreducible uncertainty, which can differ in degree for distinct random processes (their prediction can be more or less correct). If the degree of uncertainty characterizing a process approaches zero, then it becomes deterministic and exactly predictable. Consequently, the regularity of a random process can range from being completely unpredictable to being deterministic.

Suitable projection of a random trajectory along the time axis of the multidimensional behavior space yields a multivariate time series (one component series associated with each distinct axis of phase space). We use the denotations *random trajectory* and *multivariate time series* interchangeably. The concept of random trajectory or multivariate time series is very general and can accommodate nonlinearities like deterministic chaos (e.g., Casdagli, Eubank, Farmer, & Gibson, 1992) and sudden phase transitions (e.g., Molenaar & Hartelman, 1996; Molenaar & Newell, 2003). To ease the presentation, however, attention will be restricted to linear multivariate time series whose component series are real-valued and normally distributed, whereas time is considered to proceed in discrete equidistant steps.

The following notational conventions will be used: vector-valued variables are denoted by boldface lower-case letters, matrix-valued variables by boldface upper-case letters. Roman letters are used for manifest variables, Greek letters for latent variables. The symbol Σ denotes summation, where the subscript indicates the index variable. The prime symbol (′) denotes transposition.

A p-variate time series **y**(t) in discrete time t is characterized by a so-called cylinder set of finite-dimensional distributions P(**y**;t) = Prob[**y**(t) < **y**], P(**y**1, **y**2; t1, t2) = Prob[**y**(t1) < **y**1; **y**(t2) < **y**2], and so forth (cf. Brillinger, 1975). Accordingly, **y**(t) can be regarded as a random time-dependent function and we can consider its first-order moment function, second-order moment function, and so forth. In general, these moment functions can be time-varying. If, however, the first-order moment function is constant, E[**y**(t)] = **c**y, where **c**y is a constant p-variate vector, then **y**(t) is called first-order stationary. If its second-order central moment function only depends on the lag k = t2 − t1, where t1 and t2 are arbitrary time points, E[(**y**(t1) − **c**y(t1)), (**y**(t2) − **c**y(t2))′] = cov[**y**(t1), **y**(t2)′] = **C**y(k), where **C**y(k) is a (p, p)-dimensional matrix for each k, then **y**(t) is called second-order stationary. If **y**(t) is both first- and second-order stationary, then it is called weakly stationary. Notice that a weakly stationary Gaussian time series is also strongly stationary, that is, its finite-dimensional distributions (not just the first two moment functions) do not depend on time.

ERGODIC PROCESSES

As indicated earlier, an ergodic process is a process in which the structures of IEV and IAV are (asymptotically) equivalent. For nonergodic processes this equivalence does not hold. To prove that a process is ergodic, either deductively, in case the probability laws characterizing the process are known, or else inductively, can be difficult. Tong (1990) gave some detailed deductions for special processes. Only recently have useful inductive tests of ergodicity for single-participant time series become available in the econometric literature (e.g., Domowitz & El-Gamal, 2001).

The situation for Gaussian time series (our standard assumption in this article) is, however, quite simple. Consider the following result in Hannan (1970, p. 201): A Gaussian process **y**(t), t = 0, ±1, ..., is ergodic if it obeys the following restrictions: (a) it is weakly stationary (and hence strictly stationary); (b) it does not contain cyclic trends.[2] Hence a Gaussian time series is ergodic only if its first-order moment function (mean vector) and second-order moment function, lagged covariance matrices **Cy**(k), k = 0, ±1, ..., are invariant in time. The situation for non-Gaussian processes is much more complex and will not be considered here (cf. Borovkov, 1998, for general characterizations).[3]

This rather general description of ergodicity (where all technical details have been omitted; cf. e.g., Petersen, 1983, for a thorough introduction) will be illustrated shortly by applying it to the longitudinal factor model. But first the following qualification is in order. Ergodicity was introduced by Boltzmann in his attempts to provide a statistical mechanical foundation for the thermodynamics of systems in equilibrium. In this respect, ergodic theory has failed (cf. Earman & Rédei, 1996; Guttmann, 1999; Sklar, 1993). For our purposes, however, this is not relevant. We need only ergodic theory as the proper mathematical–statistical theory about the conditions under which the structures of IEV and IAV are equal. As a purely mathematical discipline, ergodic theory has thrived since it was first proposed by Boltzmann. For instance, it plays an important role in the new paradigm of nonlinear dynamics (e.g., Sinai, 1994) and the advanced theory of Markov processes (e.g., Borovkov, 1998; Keller, 1997).

Let's close this section with a specification of the conditions under which a standard longitudinal factor model of IEV is ergodic (and hence also models the

---

[2]Hannan specifies the second condition as follows: the spectral density $\mathbf{S}(\omega)$ of $\mathbf{y}(t)$ has no jumps. The spectral density $\mathbf{S}(\omega)$ of $\mathbf{y}(t)$ is defined as: $\mathbf{S}(\omega) = \mathrm{cov}[\mathbf{y}(\omega), \mathbf{y}(\omega)^*]$, where $\mathbf{y}(\omega)$ denotes the Fourier transform of $\mathbf{y}(t)$ at frequency ω, and the asterisk (*) denotes transposition in combination with conjugation. $\mathbf{S}(\omega)$ has no jumps in case $\mathbf{y}(t)$ does not have cyclic trends.

[3]For an ergodic Gaussian process, the structures of IEV and IAV are asymptotically equivalent. To further detail this relation it is necessary to introduce the concept of invariant measure. Consider an infinitely large population of participants, where the life history (trajectory in behavior space) of each participant obeys the same ergodic Gaussian process. This collection of individual trajectories allows for the definition of a (Borel) probability measure μ_t characterizing the density of trajectory values in phase space at each time t. In addition, the common dynamical law underlying the individual trajectories will be inherited by the dynamical law according to which μ_t evolves in time. We formally specify the latter as the (measurable) automorphism τ of phase space, mapping μ_t to $\tau\mu_t = \mu_{t+1}$, $t = 0, \pm1, \ldots$. If $\tau\mu = \mu$ for all t, then μ is called an invariant measure and the transformation τ is called measure-preserving. For any integrable function f we define:

$$f_{IEV} = \int_Y f(y)\mu(dy),$$

where $_Y$ is the phase space and μ is the invariant measure. In addition, for participant S belonging to the population of participants at stake we define:

$$f_{IAV} = \lim_{\tau\to\infty} \tau^{-1}\Sigma_t f(y_S(t)),$$

where $y_S(t)$, $t = 0, \pm1, \ldots$, denotes the trajectory in behavior space of participant S. For suitably chosen f, f_{IEV} captures the IEV structure of the process and f_{IAV} the (asymptotic) IAV structure. Because our stationary Gaussian process is ergodic, it follows that $f_{IEV} = f_{IAV}$.

structure of IAV). Consider T fixed time points $t = 1, \ldots, T$. In an infinitely large population of participants $i = 1, 2, \ldots$, the standard longitudinal factor model is defined as (Jöreskog, 1979)

$$\mathbf{y}_i(t) = \Lambda_t \eta_i(t) + \varepsilon_i(t),\ t = 1, \ldots, T;\ T \text{ fixed};\ i = 1, 2, \ldots,$$
$$\eta_i(t) = \mathbf{B}_{t,t-1}\eta_i(t-1) + \zeta_i(t),\ t = 2, \ldots, T,$$

where $\mathbf{y}_i(t)$ is the (centered) p-variate vector of manifest variables for participant i, Λ_t is a (p, q)-dimensional matrix of factor loadings at time t, $\eta_i(t)$ is a q-variate longitudinal latent factor at time t, $\varepsilon_i(t)$ is p-variate Gaussian measurement error at time t: $\varepsilon_i(t) \sim \aleph(\mathbf{0}, \Theta_t)$, $\mathbf{B}_{t,t-1}$ is the (q, q)-dimensional matrix of regression weights linking $\eta_i(t)$ to $\eta_i(t-1)$, and $\zeta_i(t)$ denotes q-variate Gaussian innovation at time t: $\zeta_i(t) \sim \aleph(\mathbf{0}, \Psi_t)$. It is tacitly understood throughout this article that the (p, p)-dimensional covariance matrix of measurement errors Θ is always diagonal: $\Theta = \text{diag}[\theta_1, \theta_2, \ldots, \theta_p]$.

This model is ergodic if it has the restricted form:

$$\mathbf{y}_i(t) = \Lambda\eta_i(t) + \varepsilon_i(t),\ t = 1, \ldots, T;\ T \text{ fixed};\ i = 1, 2, \ldots,$$
$$\eta_i(t) = \mathbf{B}\eta_i(t-1) + \zeta_i(t),\ t = 2, \ldots, T,$$

where Λ is invariant over time, $\varepsilon_i(t)$ has constant covariance: $\varepsilon_i(t) \sim \aleph(\mathbf{0}, \Theta)$, $\mathbf{B}$ is invariant over time, and $\zeta_i(t)$ has constant covariance: $\zeta_i(t) \sim \aleph(\mathbf{0}, \Psi)$. In addition, the absolute value (modulus) of each eigenvalue of $\mathbf{B}$ has to be strictly less than 1. In case a longitudinal factor solution does not obey the latter restricted form, one cannot validly generalize to individual (IAV) applications.

TEST THEORY AND ERGODICITY

This section is devoted to the classic book on test theory by Lord and Novick (1968), which still provides the foundational structure on which all our psychological tests and scales are based. We follow the line of thought in the beginning of that book, by means of a concise summary and selected quotes. Based on this textual evidence, I argue that test theory does not apply to nonergodic processes. Some negative consequences of this state of affairs for applied test theory are indicated, here and in the section entitled "Some Proof."

Lord and Novick (1968) defined the true score of a fixed person as the expected value of the observed score of this person with respect to his or her propensity distribution of observed scores. The latter propensity distribution is characterized as a "distribution function defined over repeated statistically independent measurements on the same person" (p. 30). It is assumed that the repeated measurements do not affect the person in that in each replication the person responds without any aftereffects of previous assessments (e.g., due to memory, habituation, etc.). With respect to this within-subject definition of true score (and of error score as the difference between observed score and true score), Lord and Novick made the following remark: "The true and error scores defined above are not those primarily considered in test theory … . They are, however, those that would be of interest to a theory that deals with individuals rather than with groups (counseling … rather than selection)" (p. 32).

In contrast, the true and error scores that *are* considered in test theory are defined in terms of between-subjects variation: "Primarily, test theory treats individual differences or, equivalently, measurements over people." (p. 32).

It is clear that the original definition of true score given by Lord and Novick (1968) applies only to IAV. The propensity distribution of a single participant is obtained by means of repeated measurements of this participant. His or her true score is defined as the mean of this propensity distribution. Then, for reasons that I have criticized elsewhere (Molenaar, 2003, chap. 3), Lord and Novick consider it practically impossible to obtain a person's propensity distribution and therefore test theory is further developed for IEV. They acknowledge the limitations of this paradigm shift, although in a somewhat oblique way. It is not clear whether their statement that the original definition of true score would be of interest to a theory that deals with individual counseling also implies its converse. Namely, that a test theory based on IEV (i.e., test theory as we know it) is not of interest to individual counseling. Yet, this is what we should conclude in case the psychological process to which test theory is applied is nonergodic. In that case the structures of IEV and IAV can differ to an arbitrary degree, up to being completely unrelated. Hence, claims based on classical test theory that a test is valid and reliable cannot be generalized to individual assessments of development, learning, or any other nonstationary process. Given that such individual assessments commonly occur in applied psychology, one can appreciate the necessity to develop a test theory according to the original true score definition of Lord and Novick (1968; i.e., a test theory based on IAV). For further elaboration of this critique of test theory, see Molenaar (2003, chap. 3).

HETEROGENEITY

Stationarity is a form of homogeneity in time that sanctions pooling across time points for the purpose of estimation in single-participant time series analysis. The assumption of homogeneity of a population of participants fulfils the same role in analyses of IEV. Alternatively, nonstationarity is a form of heterogeneity in time that complicates estimation in single-participant time series analysis (Molenaar, 1994; Priestley, 1989; cf. Molenaar & Newell, 2003, for an empirical application).

Yet the (deterministic or random) variation in time of parameters of a process, yielding nonstationarity, is not the only conceivable kind of heterogeneity in time. More drastic versions of such heterogeneity are exemplified by processes undergoing qualitative shifts in their dynamical regimes (cf. Kim & Nelson, 1999, for computational details and empirical illustrations).

Let us consider a form of heterogeneity that appears to be stronger than nonergodicity due to time-varying parameters. Suppose that, in a population of participants, the trajectory in behavior space of each single participant obeys a different factor model. Hence, each participant has his or her own personal IAV factor model, which may differ from all other participants in the number of factors, the pattern and numerical values of loadings, and/or the error variances. Although nonstationarity still implies that the same dynamical model applies to all members of a population, this homogeneity assumption at the level of dynamical regimes now is dropped. Suppose also that we randomly select N participants and register the values of their trajectories in behavior space at a fixed time t_*. Next, the data set thus obtained, $\{\mathbf{y}_i(t_*); i = 1, \ldots, N; t_* \text{ fixed}\}$, is subjected to a standard (IEV) factor analysis. Exactly this scenario has been used in a simulation study by Molenaar (1997) and it was found that a standard q-factor solution, with $q = 1$ or $q = 2$, yields satisfactorily fitting solutions that bear no resemblance to the idiosyncratic factor models describing the life history of each individual participant in the sample. Using the same scenario, equivalent results have been obtained in standard (IEV) longitudinal factor analyses of data sets $\{\mathbf{y}_i(t_k); i = 1, \ldots, N; k = 1, 2; t_k \text{ fixed}\}$ (Molenaar, 1999). The same results were again obtained in a quantitative genetical analysis of simulated heterogeneous twin data (Molenaar, Huizenga, & Nesselroade, 2002).

These simulation studies, each employing the same scenario described earlier, provide converging evidence that standard factor analysis of IEV appears to be quite insensitive to the presence of substantial heterogeneity. This is noteworthy, because it is a standard assumption of the latter factor model that its parameters are invariant (fixed) across all participants in the population. Yet, despite flagrant violation of this assumption in the behavior spaces used in the simulations, standard factor analysis of (cross-sectional, longitudinal, and twin) data sampled from these spaces yields satisfactorily fitting solutions: likelihood ratio tests are unable to reject solutions with one or two common factors, even with substantial sample sizes, and nothing in the obtained solutions (Lagrange multiplier tests, standard errors, etc.) is indicative of the heterogeneity present in the population. It appears that only dedicated factor analysis of IAV can uncover this.

DEVELOPMENT IMPLIES HETEROGENEITY

We concisely describe some strong arguments why one should expect there to be substantial heterogeneity of the sort described in the previous section. This section is based on Molenaar, Boomsma, and Dolan (1993) and Molenaar and Raijmakers (1999), to which the reader is referred for additional elaborations.

If one compares the maximum amount of information that can be stored genetically with the information required for the growth of the embryonic brain, then it turns out that these amounts differ among several orders of magnitude. Even if (counterfactually) all genetically stored information would be used to specify the intricate wiring of neurons in a developing brain, this would be far too little. Hence brain development can be successful only if the developmental processes concerned are self-organizing. Self-organization is a characteristic of nonlinear dynamical processes and current mathematical–biological models of self-organizing growth are of this kind (e.g., nonlinear reaction–diffusion models of biological pattern formation; cf. Kauffman, 1993; Möller & Swaddle, 1997; Murray, 1993). It is an important feature of self-organizing growth that it is ordered, but not invariant. Edelman (1987) presented impressive empirical evidence for the existence of variability of the detailed wiring in homologous neural structures within the same organism or between genetically identical organisms reared in standardized environments (cf. Molenaar et al., 1993, for additional results).

Given that the detailed wiring of neural networks in the brain of even genetically identical organisms shows considerable variation, and given that the brain is causally related to the stream of behavior, it can be expected that this variability is inherited by the structure of psychological processes. This implies that these psychological processes are heterogeneous. Of course, several alternative scenarios leading up to heterogeneity can be given, for instance the multitude of environmental effects impinging on diverse levels of a growing, developing, adapting, and accommodating participant, but the argument given earlier should suffice for our present purposes.

SOME PROOF

Simulation is a powerful investigative tool. For instance, it is indispensable for the study of small sample properties. In the section on heterogeneity we saw that it can provide preliminary answers to new questions. Ideally, however, such answers should be underpinned by proof. We outline some proof of the insensitivity of IEV factor analysis to the presence of heterogeneity, our main result obtained in the simulation studies concerned. What follows is based on Kelderman and Molenaar (2003).

To get started we make several simplifications. Consider again a population of participants, where the trajectory in behavior space of each participant obeys a person-specific (IAV) one-factor model. So the possibility that participants can differ with respect to the number of common factors no longer obtains. In fact, we assume that the personal one-factor models of participants can differ only in the numerical values of the factor loadings. At an arbitrary, but fixed, time t_*, we then have (remember that $\mathbf{y}_i(t_*)$ is centered):

$$\mathbf{y}_i(t_*) = \lambda_i \eta_i(t_*) + \varepsilon_i(t_*),\ i = 1, 2, \ldots;\ t_* \text{ fixed.}$$

where λ_i denotes the p-variate vector of individual factor loadings for participant i, $\eta_i(t_*)$ is the factor score for participant i, and $\varepsilon_i(t_*)$ is the measurement error for this participant. Because t_* is arbitrary and fixed, we can simplify notation:

$$\mathbf{y}_i = \lambda_i \eta_i + \varepsilon_i,\ i = 1, 2, \ldots.$$

Clearly, the latter model equation has the form of a simple one-factor model. However, the factor loadings in λ_i are random instead of fixed. Suppose that λ_i follows a p-variate Gaussian distribution in the population of participants, with mean vector υ and (p, p)-dimensional covariance matrix Ξ: $\lambda_i \sim \aleph(\upsilon, \Xi)$. Simplifying further (cf. Kelderman & Molenaar, 2003, for the general case), it is assumed that Ξ is a (p, p)-variate diagonal covariance matrix: $\Xi = \text{diag}[\xi_1, \xi_2, \ldots, \xi p]$. In addition the usual assumptions for the standard one-factor model are added: $\eta_i \sim \aleph(0, \psi)$, $\varepsilon_i \sim \aleph(\mathbf{0}, \Theta)$, whereas η_i is uncorrelated with ε_i. To this we also add the assumption that λ_i is uncorrelated with ε_i and η_i. It then can be proved by straightforward expansion in terms of moments up to fourth order that the communal part of our factor model with random loadings is:

$$\text{cov}[\lambda_i \eta_i, (\lambda_i \eta_i)'] = E[\lambda_i]\text{var}[\eta_i]E[\lambda_i]' = \upsilon\psi\upsilon'$$

which is again the communal part of a standard (IEV) one-factor model with fixed loadings υ.

The conditions under which this result is obtained are strong. The assumption that error variances are homogeneous across participants can easily be dropped and changed into: $\varepsilon_i \sim \aleph(\mathbf{0}, \Theta_i)$; that is, each participant i has measurement errors with idiosyncratic covariance matrix Θ_i. In case $\mathbf{y}_i$ is not centered and $E[\eta_i] \neq 0$, the same result is obtained: the communal part of the latter one-factor model with random loadings and nonzero factor mean is also indistinguishable from the communal part of a standard one-factor model with fixed loadings. Kelderman and Molenaar (2003) presented evidence from a Monte Carlo study, showing that the likelihood ratio associated with the one-factor model with random loadings has the regular chi-square distribution with nominal degrees of freedom. Generalization of these results to q-factor models (including longitudinal factor models) with random loadings is expected to be straightforward (but has not been undertaken yet).

Even the limited proof outlined here can be seen to have important consequences. The main cause of this is the fact that the proven structural equivalence of the one-factor model with random loadings and the standard one-factor model with fixed loadings does not depend on the variances of the random loadings $var[\lambda_i] = \Xi = diag[\xi_1, \xi_2, \ldots, \xi_p]$. Suppose that a standard one-factor model is found to yield a satisfactory fit in some applied setting, for instance, in test construction according to classical test theory (cf. Lord & Novick, 1968, section 24.3). Our proof shows that the finding that a standard one-factor model explains the structure of IEV leaves entirely open the possibility that the actual factor loadings λ_i, characterizing the IAV of each individual participant i, differ to arbitrary degrees from the fixed loadings in the standard solution. Depending on the variances $diag[\xi_1, \xi_2, \ldots, \xi_p]$ of the random loadings, it can only be stated that the probability is .95 that the participants in the population have trajectories in behavior space obeying one-factor models of which the loadings lie in the interval $\upsilon \pm 1.96\,\xi$, where the p-variate vector ξ is defined as $\xi = [\sqrt{\xi_1}, \sqrt{\xi_2}, \ldots, \sqrt{\xi_p}]'$. But ξ is unknown in standard factor analysis, and for all we know the entries of ξ can have any nonnegative value. If one would use the standard one-factor solution to conclude that a test has high reliability, then this is compatible with an actual state of affairs in which the test has almost zero reliability in an infinitely large subset of participants in the population. If one would estimate (predict) the factor scores in the standard factor model, then the correlation with the true factor scores in the model with random loadings may even become negative (as has been found in the simulation study reported in Molenaar, 1999). Obviously, this will jeopardize the confidence one can have in decisions, arrived at in individual counseling based on the (IEV) factor scores obtained with a nominally reliable test.

A CAVEAT

We happen to live in an era in which multilevel modeling appears to be rather popular. In Molenaar (2003, chap. 2), the vices and virtues of multilevel modeling, more specifically of the so-called latent growth curve model, have been discussed at length. In addition, it is proved that the latent growth curve model is a special case of the latent simplex model (which for some reason appears to be less popular). Here, however, we address the latent growth curve model, a model of the structure of IEV, to make explicit the difference with the factor model of IAV with random loadings that figured in the previous section. For this purpose we only need the simplest kind of latent growth curve model (see Molenaar, 2003, for the general case).

Let $t = 1, \ldots, T$ denote a sequence of fixed equidistant time points (T usually is small; e.g., $T = 5$). Let $y_i(t)$, $t = 0, \ldots, T$; $i = 1, 2, \ldots$, denote univariate repeated measurements obeying the following model (without loss of generality $y_i(t)$ is assumed to be centered):

$$y_i(t) = \alpha_i + \beta_i t + \varepsilon_i(t),\ t \in \{0, \ldots, T\};\ T \text{ fixed.}$$

It is immediately apparent (and well-known) that this model can be rewritten as a constrained two-factor model. Let the common latent factor be $\eta_i = [\alpha_i, \beta_i]'$, fix the first column of the (T + 1, 2)- dimensional matrix of loadings Λ at 1 and the second column at t = 0, ..., T.

Because the latent growth curve model is equivalent to a constrained standard factor model, it is vulnerable to the same limitations as discussed in the previous sections. Perhaps the distance between the latent growth curve model and the factor model with random loadings is even more extreme than for the standard factor model, because in the standard factor model the loadings are fixed and latent, whereas in the latent growth curve model they are fixed and manifest (i.e., have known values).

AN EMPIRICAL COMPARISON OF THE STRUCTURES OF IEV AND IAV

Borkenau and Ostendorf (1998) described an interesting replicated time series experiment in which they obtained 30-variate time series of daily scores on a personality test. There were 22 participants participating in this experiment, who were each measured on 90 consecutive days. The test presumably measures five latent personality dimensions (IEV factors): Neuroticism, Extraversion, Agreeableness, Conscientiousness, and Intellect. Borkenau kindly provided the raw data (no missing values!) for additional analysis.

This data set allows for a direct comparison of the structures of IEV and IAV. To determine the structure of IEV, a (maximum likelihood) confirmatory oblique factor analysis was carried out on a robust estimate of the (30, 30)-dimensional correlation matrix. The normative loading pattern of the 30 items on the five factors was almost completely recovered.

The 30-variate time series of each individual was also subjected to advanced time series analysis to determine the best fitting factor model for the structure of IAV. We then obtain results that are reminiscent of the scenario discussed earlier in the section on heterogeneity: each individual has his or her own personal factor model. The time series of some participants obey a two-factor model, of others a three-factor model, and again others a four-factor model. Those participants whose time series obeys a factor model with the same number of factors differ in various other respects, such as loading pattern and values, measurement error variances, and/or the process model according to which a participant's latent factor scores vary in time.

Clearly, in this particular application the obtained structures of IEV and IAV differ substantially. In addition, the behavior space occupied by the 22 participants appears to be quite heterogeneous. If one would use the nominal five-factor IEV structure to predict an individual's development, whereas the IAV structure of this particular participant obeys a three-factor model, say, then it can be expected that prediction success will be dubious. A complete description of this application is given in Hamaker and Molenaar (2004), together with a hybrid model linking IEV and IAV structures.

CONCLUSION

The classical ergodic theorems show that the structures of IEV and IAV are equivalent only under very stringent conditions. Only for the special case of Gaussian processes do these conditions boil down to the requirement of stationarity of the mean and covariance function. Almost by definition habituation, learning, and development are nonstationary, and hence nonergodic (for non-Gaussian processes the equivalence of [non]stationarity and [non]ergodicity has exceptions in both directions). Even a few seconds of electro-encephalographic registration, obtained under stationary conditions with a resting participant with eyes closed, is nonstationary. For nonergodic processes there is no scientifically respectable alternative but to study the structures of IAV and IEV for their own sakes. For further information about investigation of the structures of IAV under realistic conditions, including alternative perspectives on generalization, the reader is referred to Nesselroade and Molenaar (1999) and Hamaker, Dolan, and Molenaar (2003).

One can conceive of nonstationarity/nonergodicity as a rather mild form of heterogeneity in time. Both empirical and mathematical evidence obtained in the biological sciences is indicative of the presence of stronger forms of structural heterogeneity. Converging evidence obtained in simulation studies, as well as analytic proof in a special case, shows that it is difficult (up to impossible, in the special case) to detect such stronger forms of heterogeneity in standard factor analysis of IEV. Some disturbing implications of this state of affairs for test theory were considered.

The key to the dedicated study of structures of IAV is (replicated) time series analysis. At the University of Amsterdam we developed sophisticated time series analysis software for most conceivable cases. Following the classification of latent variable models in Bartholomew (1987), special freeware is available for the case in which manifest and latent processes are both metrical (state–space modeling; cf. Molenaar, 1985), for the case in which the manifest process is categorical and the latent process is metrical (generalized linear dynamical modeling; cf. Fahrmeir & Tutz, 1994), and for the case in which the manifest process is metrical or categorical and the latent process is categorical (hidden Markov modeling; cf. Elliott, Aggoun, & Moore, 1995). Hamaker et al. (2003; cf. also Nesselroade & Molenaar, 1999) discussed analysis of replicated time series when the number of participants (replicates) and repeated observations within participants (time points) is relatively small.

In the course of the discussion we introduced the concept of a behavior space. Cross-sections of this behavior space at one or more fixed time points yield data amenable to analyses of IEV, whereas time-slices of the trajectory (life history) of one or more participants yield data amenable to analyses of IAV. The concept of behavior space is reminiscent of the concept of an ensemble in statistical mechanics (cf. MacKey, 1993, for distinctive definitions in the senses of Boltzman and of Gibbs). The concept of behavior space has been instrumental in conveying the basic tenets of this article. Additional arguments for its usefulness are given in Molenaar (2003, chap. 3).

The unavoidable consequence of the ergodic theorems is that psychometrics and statistical modeling as we now know it in psychology are incomplete. What is lacking is the scientific study of the individual, his or her structure of IAV, for its own sake. Scientific psychology can only become complete if it includes the idiographic point of view, alongside the nomothetic point of view. In fact, replicated time series designs and analyses allow for the induction of nomothetic laws for idiographic patterns of IAV. The idiographic approach, firmly based on adequate time series analysis, should cover all main areas in psychology, including cognitive science, clinical psychology, and social psychology (e.g., small groups dynamics), to mention a few. Generalized linear dynamical modeling opens up the possibility to fit dynamical versions of item-response models to single-participant time series obtained in educational psychology.

Coming at the end of this manifesto, I would like to make a prediction about the future trajectory of psychology itself. The total number of persons that will have lived on the earth before it becomes inhospitable to life (heat death) is finite (about 10^{14}, despite our assumptions about infinitely large populations). For a modern physicist this number is not really large (Avogadro's number of molecules in a mole of gas is about 10^{24}). Given the ongoing spectacular increase in computational capacity and facilities (e.g., quantum computing at subatomic levels), and given also the expected increase in possibilities for on-line sensoring (e.g., micro-sensors in cloths), it can be predicted that in the near future it will be possible to continuously assess and store the unfolding life history (trajectory in behavior space) of each individual. Added to this is detailed knowledge of the genotype of each individual and other relevant background variables. Using recursive optimal control techniques, it then becomes possible to personalize treatments in place and time (cf. Molenaar, 1987, for an application of on-line optimal control in an individual psychotherapeutic process). The more we approach this state of affairs, the more the power of the idiographic approach as described here will become evident.

ACKNOWLEDGMENT

This article is an extended version of an invited talk held at the 13th International Meeting and the 68th Annual Meeting of the Psychometric Society, Cagliari, Italy, July 7–10, 2003.

REFERENCES

Bartholomew, D. J. (1987). *Latent variable models and factor analysis.* London: Charles Griffin & Company.

Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Personality Research, 32,* 202–221.

Borovkov, A. A. (1998). *Ergodicity and stability of stochastic processes.* New York: Wiley.

Brillinger, D. R. (1975). *Time series: Data analysis and theory.* New York: Holt, Rinehart & Winston.

Casdagli, M., Eubank, S., Farmer, J. D., & Gibson, J. (1992). State–space reconstruction in the presence of noise. In D. K. Campbell, R. E. Ecke, & J. M. Hyman (Eds.), *Nonlinear science: The next decade* (pp. 52–98). Cambridge, MA: MIT Press.

de Groot, A. D. (1954). Scientific personality diagnosis. *Acta Psychologica, 10,* 220–241.

Domowitz, I., & El-Gamal, M. A. (2001). A consistent nonparametric test of ergodicity for time series with applications. *Journal of Econometrics, 102,* 2001, 365–398.

Earman, J., & Rédei, M. (1996). Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science, 47,* 63–78.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* New York: Basic.

Elliott, R. J., Aggoun, L., & Moore, J. B. (1995). *Hidden Markov models: Estimation and control.* New York: Springer-Verlag.

Fahrmeir, L., & Tutz, G. (1994). *Multivariate statistical modeling based on generalized linear models.* Berlin: Springer-Verlag.

Guttmann, Y. M. (1999). *The concept of probability in statistical physics.* Cambridge, England: Cambridge University Press.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points *T* exceeds the number of cases *N:* Raw data likelihood. *Structural Equation Modeling, 10,* 352–379.

Hamaker, E. L., & Molenaar, P. C. M. (2004). The integrated trait-state model. Manuscript submitted for publication.

Hannan, E. J. (1970). *Multiple time series.* New York: Wiley.

Jöreskog, K. G. (1979). Statistical estimation of structural equation models in longitudinal developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–351). New York: Academic.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution.* Oxford, England: Oxford University Press.

Kelderman, H., & Molenaar, P. C. M. (2003). *On individual differences in factor structure and the fit of the standard 1-factor model.* Manuscript submitted for publication.

Keller, G. (1997). *Equilibrium states in ergodic theory.* Cambridge, England: Cambridge University Press.

Kim, C. J., & Nelson, C. R. (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications.* Cambridge, MA: MIT Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

MacKey, M. C. (1993). *Time's arrow: The origins of thermodynamic behavior.* Berlin: Springer-Verlag.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50,* 181–202.

Molenaar, P. C. M. (1987). Dynamic assessment and adaptive optimization of the therapeutic process. *Behavioral Assessment, 9,* 389–416.

Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. C. Clogg (Eds.), *Analysis of latent variables in developmental research* (pp. 155–180). Newbury Park, CA: Sage.

Molenaar, P. C. M. (1997). Time series analysis and its relationship with longitudinal analysis. *International Journal of Sports Medicine, 18,* 232–237.

Molenaar, P. C. M. (1999). Longitudinal analysis. In H. J. Ader & G. J. Mellenbergh (Eds.), *Research methodology in the social, behavioral and life sciences* (pp. 143–167). London: Sage.

Molenaar, P. C. M. (2003). *State space techniques in structural equation modeling: Transformation of latent variables in and out of latent variable models.* Available from: http://www.fmg. uva.nl/ml/ publicaties_ml.cfm

Molenaar, P. C. M., Boomsma, D. I., & Dolan, C.V. (1993). A third source of developmental differences. *Behavior Genetics, 23,* 519–524.

Molenaar, P. C. M., & Hartelman, P. (1996). Catastrophe theory of stage transitions in metrical and discrete stochastic systems. In A. von Eye & C. C. Clogg (Eds.), *Categorical variables in developmental research: Methods of analysis* (pp. 107–130). San Diego: Academic.

Molenaar, P. C. M., & Raijmakers, M. E. J. (1999). Additional aspects of third source variation for the genetic analysis of human development and behavior. *Twin Research, 2,* 49–52.

Molenaar, P. C. M., Huizenga, H. M., & Nesselroade, J. R. (2002). The relationship between the structure of inter-individual and intraindividual variability: A theoretical and empirical vindication of Developmental Systems Theory. In U.M. Staudinger & U. Lindenberger (Eds.), *Understanding human development* (pp. 339–360.). Dordrecht, The Netherlands: Kluwer.

Molenaar, P. C. M., & Newell, K. M. (2003). Direct fit of a theoretical model of phase transition in oscillatory finger motion. *British Journal of Mathematical and Statistical Psychology, 56,* 199–214.

Möller, A. P., & Swaddle, J. P. (1997). *Asymmetry, developmental stability, and evolution.* Oxford, England: Oxford University Press.

Murray, J. D. (1993). *Mathematical biology* (2nd ed.). Berlin: Springer-Verlag.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223–250). Thousand Oaks, CA: Sage.

Petersen, K. (1983). *Ergodic theory.* Cambridge, England: Cambridge University Press.

Priestley, M. B. (1989). *Nonlinear and nonstationary time series analysis.* London: Academic.

Sinai, Y. G. (1994). *Topics in ergodic theory.* Princeton, NJ: Princeton University Press.

Sklar, L. (1993). *Physics and chance: Philosophical issues in the foundations of statistical mechanics.* Cambridge, England: Cambridge University Press.

Tong, H. (1990). *Non-linear time series: A dynamical system approach.* Oxford, England: Oxford University Press.