Chapter 4
The Two Disciplines of Scientific Psychology, or: The Disunity of Psychology as a Working Hypothesis

Denny Borsboom, Rogier A. Kievit, Daniel Cervone and S. Brian Hood

Anybody who has some familiarity with the research literature in scientific psychology has probably thought, at one time or another, 'Well, all these means and correlations are very interesting, but what do they have to do with me, as an individual person?'. The question, innocuous as it may seem, is a deep and complicated one. In contrast to the natural sciences, where researchers can safely assume that, say, all electrons are exchangeable save properties such as location and momentum, people differ from each other. Furthermore, it is not obvious that these differences can be treated as irrelevant to the structure of the organisms in question, i.e., it is not clear that they can be treated as 'noise' or 'error'. The problem permeates virtually every subdiscipline of psychology, and in fact may be one of the reasons that progress in psychology has been limited. As Lykken (1991, pp. 3–4) hypothesizes:

> Psychology isn't doing very well as a scientific discipline and something seems to be wrong somewhere. This is due partly to the fact that psychology is simply harder than physics or chemistry, and for a variety of reasons. One interesting reason is that people differ structurally from each other and therefore cannot be understood in terms of the same theory since theories are guesses about structure.

Lykken's hypothesis—that the lawfulness in human behavior, and whatever underlies it, may be person-specific—has potentially far-reaching consequences. Taken to its limit, the truth of the hypothesis would imply that scientific psychology would involve the construction of theories of human behavior on a case-by-case basis—an unmanageable task. In addition, it is not clear whether such an approach would not be contrary to scientific practice as we currently know it, which seeks to generalize theories over the objects that they apply to. It is hard, for instance, to imagine a physics that involves constructing a new theory of free fall for every piece of rock we may want to study. Nevertheless, the processes that underlie your behavior are probably more complicated than, say, the gravitational dynamics that

D. Borsboom (✉)
Department of Psychology
University of Amsterdam
1012 ZA Amsterdam, The Netherlands
e-mail: d.borsboom@uva.nl

J. Valsiner et al. (eds.), *Dynamic Process Methodology in the Social and Developmental Sciences,* 67
DOI 10.1007/978-0-387-95922-1_4, © Springer Science+Business Media LLC 2009

underlie the movements of planets in the solar system, and hence Lykken's hypothesis has some initial plausibility.

Given the magnitude of the problems involved in constructing person-specific theories and models, let alone in testing them, it is not surprising that scholars have sought to integrate inter-individual differences and intra-individual dynamics in a systematic way. This may involve, for instance, constructing theories that apply to subgroups of people who are homogeneous at the relevant level of the processes under study. In such a case, full generalizability of theories to individuals may not be possible, but it would be possible to give a systematic account of how inter-individual differences in intra-individual processes are distributed in the general population, and how they arise in human development. This would render the task of partially homogenizing people, by allocating them to homogeneous subgroups, at least somewhat manageable.

The call for integration of research traditions dates back at least to Cronbach's (1957) lament of the disintegrated state of scientific psychology as it existed in the 1950s. In this paper, Cronbach (1957) sketched what he viewed as a solution to the problem of integrating both research on inter-individual differences (which he identified with 'correlational psychology') and intra-individual processes ('experimental psychology', in his parlance):

> Correlational psychology studies only variance among organisms; experimental psychology studies only variance among treatments. A united discipline will study both of these, but it will also be concerned with the otherwise neglected interactions between organismic and treatment variables (…). Our job is to invent constructs and to form a network of laws which permits prediction. From observations we must infer a psychological description of the situation and of the present state of the organism. Our laws should permit us to predict, from this description, the behavior of organism-in-situation. (Cronbach, 1957, pp. 681–682)

One of the notable features of the scientific developments since the 1950s is that Cronbach's vision of a unified psychology has failed to materialize. Although his call for integration has been echoed by later writers who noted the gulf between the experimental and correlational styles of research and the corresponding fractionalization of scientific psychology (e.g., Sternberg & Grigorenko, 2001; Borsboom, Mellenbergh, & Van Heerden, 2003), experimental and correlational psychology have not moved much closer since 1957. Certainly, both have expanded and progressed considerably—but rarely in each other's direction; and the theories used in each of the scientific frameworks show few signs of converging into a unified system.

The fact that no integrated discipline of psychology has heretofore materialized may be related to Lykken's (1991) hypothesis of person-specific structure; for it is likely that the integration of the different schools would have been an accomplished fact, if people *were* homogeneous in the dynamic structure of their mental life and behavior. Thus, the lack of integration of research traditions invites a systematic analysis of the way that psychology treats the individual. This, then, defines the main topic of the present chapter: How does psychology treat the individual person, and which theoretical and methodological problems emerge from that treatment?

Why have research traditions on intra-individual and inter-individual differences not converged to a greater degree?

The structure of this chapter is as follows. First, we will sketch, roughly, what we perceive to be the ruling research paradigms in psychology: experimental and correlational methodology. Second, we will discuss recent methodological research into homogeneity conditions and show how their violations may affect the conclusions that researchers draw from their observations. Some particularly problematic fields are discussed in detail by focusing on the fields of intelligence and personality research. Third, we discuss possible loci of homogeneity in scientific models, and sketch the prospects for scientific psychology that may arise from these.

Ruling Paradigms

Not much has changed in the basic divisions in scientific psychology since Cronbach (1957) wrote his presidential address. True, today we have mediation and moderation analyses, which attempt to integrate inter-individual differences and intra-individual process, and in addition are able to formulate random effects models that to some extent incorporate inter-individual differences in an experimental context; but by and large research designs are characterized by a primary focus on the effects of experimental manipulations or on the structure associations of inter-individual differences, just as was the case in 1957. The rough structure of these methodological orientations is as follows.

Experimental Research

In experimental research, the researcher typically hopes to demonstrate the existence of causal effects of experimental manipulations (which typically form the levels of the 'independent variable') on a set of properties which are treated as dependent on the manipulations (their levels form the 'dependent variable'). As an example, Bargh, Chen, and Burrows (1996) created an experimental condition in which subjects were primed by words like 'bingo', 'Florida', 'wrinkle' and other words associated with the elderly, and a control condition in which they were primed with neutral words. They then measured the time it took subjects to walk from the experimental room. Bargh et al. (1996, p. 237) claim that '[p]articipants in the elderly priming condition ($M = 8.28$ s) had a slower walking speed compared to participants in the neutral priming condition ($M = 7.30$ s), $t\ (28) = 2.86$, $p < 01$, as predicted.'

One interesting and very general fact about experimental research is that such claims are never literally true. The literal reading of conclusions like Bargh et al., very prevalent among untrained readers of scientific work, is that all participants in the experimental condition were slower than all those in the control condition.

But that, of course, is incorrect—otherwise there would be no need for the statistics. As Lamiell (1987) has argued, the statements that follow from the statistical analysis (assuming the validity of the experiment and dismissing the possibility of a Type 1 error or fluke) are true 'of the average' but not 'in general' (i.e., they are true of aggregate statistics, but not true for each individual). In Bargh et al.'s (1996) research, for instance, we can be certain that some people in the experimental condition were faster than some people in the control condition (unfortunately it is hard to tell how many, as the Bargh et al. (1996) paper gives no idea of shape of the distribution of walking times, not even rough descriptives like standard deviations).

Of course, this is an entirely unsurprising fact for those acquainted with experimental research. In fact, it is so unsurprising that few researchers find it significant at all. After all, the difference between the means is in the 'right' direction, and that, for the typical researcher, is what really matters. However, the question is: in what sense is this direction the *right* direction?

In the minds of Bargh et al. (1996)—and many other experimental psychologists—the direction appears to be 'right' in the sense that it gives evidence in support of a universal law or mechanism. For instance, Bargh et al. (1996, p. 242) conclude: '[The experiments] showed that traitlike behavior is (…) produced via automatic stereotype activation if that trait participates in the stereotype.' This obviously is not intended to hold for, say, 56.7% of the people. This is supposed to be a universal law. In this respect, Bargh et al.'s research is paradigmatic for experimental research in psychology.

Clearly, the universal law is not very universal here—otherwise no t-tests would have been performed. So, there exist differences between individuals that are not attributable to the experimental manipulation. In the research tradition of experimental psychology, however, these differences are analyzed—both conceptually and statistically—as noise. The investigator 'sees' the universal mechanisms through the 'lens' of a statistical analysis, which is assumed to pick up such mechanisms. The underlying picture here is that each and every individual is an instantiation of a universal process that is uncovered by the experiment, much like mean differences in growth of crop are assumed to reflect the effects of different fertilizers (not coincidentally, the experimental design for which R.A. Fisher invented the analysis of variance). Hence, inter-individual differences are viewed as noise.

How does the individual person fit in this scheme of thinking? It appears that, within standard experimental research, the individual figures as an entity that is fully exchangeable with any other entity of the same type. This is true across subfields of psychology. Even in social psychology, a discipline that might have been expected historically to have attended to individuals' distinctive personal and socio-cultural background, individuals primarily have been conceived merely as "members of hypothetical statistical populations" (Danziger, 2000, p. 344). They thus are interchangeable elements of groups defined in terms of the experimental manipulation. The mechanisms underlying any experimental effects (apart from the inevitable 'noise') are then assumed to be homogeneous; 'the same type' is the most general type available in psychological research, namely, the human being. In research designs that allow for differences between groups of people (e.g., when a variable moderates effects) of that correct for such differences (e.g., through matching or analysis of covariance), homogeneity is required for the subgroups of people who have equivalent positions on the variables that are used for moderation analyses, matching, or analysis of covariance.

Correlational Research

One man's trash is another man's treasure. What the experimental psychologist views as error, and tries to block in all possible ways from confounding the experimental effects, is the object of study for the correlational psychologist. In correlational research, the focus is on the structure of association between variables on which people differ. Typical research findings from correlational studies are, for instance, 'people with bigger brains have higher average IQ-scores', 'extraverts do better in sales', 'there is high co-morbidity between depression and generalized anxiety', or '80% of inter-individual differences in bodily height caused by genetic differences between people'.

Such statements concern facts about inter-individual differences. It is tempting, however, to conclude that they also have meaning for a single individual. This is not generally true. To illustrate this, it is useful to use an approach to meaning in which the meaning of a statement is analyzed in terms the conditions that would render it true. As an example, the statement 'No Ravens are white' is true in all situations in which there are Ravens and none of them is white. Notice that there are various situations, e.g., involving black, blue or green ravens, which all conform to the statement above and therefore fulfill its truth conditions. Analogously, one might concoct the set of all possible situations, call it S, that would yield a heritability coefficient of 0.80 in the population, and say: 'this is what my statement means; to say that 80% of the observed variance is due to genetic variance is to say that one of the situations in S obtains'. Now, it is clear that all the situations in S involve a population of that consists of people who differ from each other. It is also clear that none of the situations in S is a situation where there are no differences between people. By extension, there is no situation in S in which there is only one individual, say, you. Thus, the statement is literally meaningless, in the sense that it has *no* truth conditions, when interpreted at the level of an individual person.

So, for instance, if you are two meters tall, the above statement about heritability does not entail that 1.80 m of that length are due to your genes and the rest to the environment. The heritability estimate is a function of variance (in this case the ratio of genetic to total variance) and that variance is, in your case, zero. So, should the rest of humanity suddenly decease from a sudden epidemic, leaving you to be the only survivor, then there would no longer be a heritability of height, because there is no variance left to define it on or estimate it from. The same holds for all correlations that are defined on inter-individual differences, except when very stringent conditions are met (to be described below).

Thus, although in some cases correlational research may yield clues to suggest the presence of universal processes, in general the results cannot be interpreted in such a way. Hence, in Lamiell's (1987) terms, results from this line of research are not true 'in general' either. However, neither are they simply true 'of the average' as the facts from experimental research may be (if it is indeed the case that the underlying mechanisms are universal and all the variance unaccounted for is noise). That is, in the case of experimental research, the facts yielded may be true of the average without any inter-individual differences that exist in the working of the mechanisms studied. This is not generally the case for correlational research. For instance, in the correlational case, full homogeneity of the studied population would consistently yield null results for the study of inter-individual differences (as these are pure noise). Thus, rather than being 'true of the average', conclusions drawn from correlational research are 'true of the inter-individual differences', and without such inter-individual differences, they have no meaning.

What does this mean for the conceptualization of the individual in correlational designs? As was argued above, in much experimental research a person is seen as the instantiation of a universal process (plus or minus error), which is varied by the experimental manipulation. In correlational research, the person functions as the instantiation of a class of people with a given position of an inter-individual differences variable (say, 'all people who are two meters tall'). Thus, the function of the individual in experimental and correlational studies is almost orthogonal. Experimental studies assume, typically, that a person does not differ from other people in relevant ways, and analyze any remaining variance as noise. Correlational studies assume, typically, that a person does differ from other people and work exactly on these differences.

Relations Between the Approaches

In general, facts from correlational research do not generalize to experimental research or vice versa. For instance, if it is true that there is a universal influence (intraindividual processes) of stereotype primes on walking speed, then this does not imply that interinter-individual differences in walking speed are correlated with the extent to which people have been primed with 'Florida'. Conversely, if it is true that inter-individual differences in normal walking speed are positively correlated with bodily height, this does not mean that surgically increasing your height will make you faster, or that walking faster will make you taller. Indeed, relations between variables may be in opposite direction in experimental versus correlations studies, without any contradiction. As an example, it may be universally true that drinking coffee increases one's level of neuroticism; then it may still be the case that people who drink more coffee are less neurotic, as illustrated in Fig. 4.1.

As can be seen from the figure, the lack of correspondence between intraindividual and interindividual relations between variables is a subgroup problem; the relation between coffee consumption and neuroticism is positive in each individual,

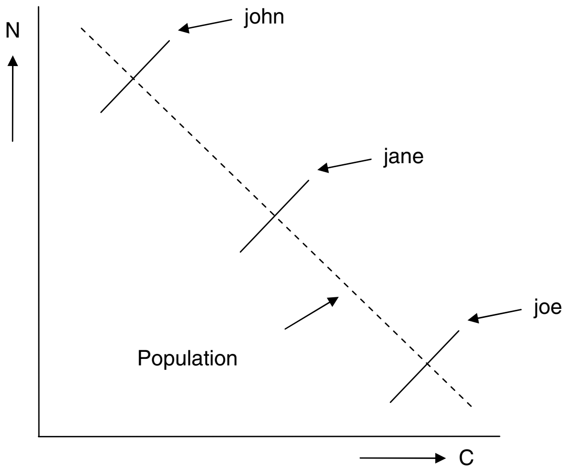

Fig. 4.1 Hypothetical relation between coffee consumption and neuroticism. For each individual, the correlation between these variables is positive, but in the population the correlation is negative

but those individuals who drink more coffee are generally less neurotic (this is, by the way, a special case of Simpson's paradox; see Simpson, 1951). As a result, the idea that correlational and experimental research can 'converge', in the sense that they render support for the *same* hypothesis—commonly viewed as a desideratum in psychological research—only makes sense in a limited set of situations—namely those in which the inter-individual differences found in correlational research are *exclusively* the result of the intraindividual processes studied in the corresponding experimental research. In situations where this is not true, it is unclear whether correlational research can 'support' the kind of hypotheses that are tested in experimental research, because these involve universal processes rather than inter-individual differences; and the set of situations in which laws concerning universal processes yield *any* predictions about the structure of inter-individual differences is highly limited.

The Role of Temporal Dynamics

The contrasting effects that may be found in correlational versus experimental designs can be disentangled if it is possible to use temporal information. For instance, intraindividual designs, that sample from the time-domain, would conceivably allow the researcher to see that something like Fig. 4.1 is indeed going on (Hamaker, Nesselroade, & Molenaar, 2007; Timmerman, Ceulemans, Lichtwarck-Aschoff, & Vansteelandt, 2009—this book). The researcher would find, in that case, that all intra-individual relations are negative, while all inter-individual relations are positive. Using within-subject experimental designs allows one to extend such analyses to experimental manipulations, thereby getting a handle on the relations that exist between intra-individual processes and inter-individual differences.

In order to gauge the possible outcomes of such research, without actually doing it, one can also use theoretical analyses of how temporal dynamics may relate to inter-individual differences and responses to experimental manipulations. This is useful because it allows for a general assessment of the structure of these relations, for instance, it allows one to assess under which circumstances results from a given research designs may be unproblematically generalized to other domains. Below we will execute such a theoretical analysis with respect to the measurement problem, by assessing the relations between person-specific measurement structures and models for inter-individual differences.

The Psychometric View: Measurement Models and Local Homogeneity

In the overwhelming majority of cases in psychology, the intended interpretations of research data go beyond the actual observations. So, for instance, researchers study IQ-scores, but want to draw conclusions about intelligence; they get observations on the reported frequency of alcohol abuse, but want say something about addiction; they get data from diagnostic interviews, but want to make inferences regarding depression. The tradition of scientific psychology is to view such observed scores as ‘indicators’ of an underlying structure (called a ‘psychological attribute’ or, somewhat misleadingly, a ‘construct’) that is measured through the indicators. Naturally, in order to gauge whether bridging the gap between intra-individual process research and interindividual research is at all possible, one requires some understanding of the relation between the measurement structures that may arise in each of these domains.

Measurement models, as they are currently used in psychology, conceptualize measurement in keeping with the idea that there exists a causal relation between the attribute measured (say, general intelligence) and the measurement outcomes (IQ-scores), in such a way that the scores causally depend on the attribute measured (Borsboom, 2005, 2008). This is most obvious in situations where models with multiple indicators are used (e.g., factor models or item response theory models). In these situations, the measurement model is formally indistinguishable from a common cause model; the latent variable (a formal stand-in for the attribute measured) functions as the common cause of the indicators. Thus, for instance, the measurement model says that the probability of developing a given depression symptom (lack of sleep, depressed mood, suicidal ideation) is a monotonically increasing function of the level of depression. Moreover, most measurement models require that, given a position on the latent variable, there are no correlations between the indicators. Thus, in this example, the level of depression ‘screens off’ these correlations. The ‘screening off’ relation is one of the defining features of a common cause model (Pearl, 2000; in the latent variable modeling literature, this property is called ‘local independence’).

One can set up measurement models both for intra-individual differences as they extend over time, and for inter-individual differences as they extend over persons. In the first situation, one typically studies one person (or a small group) by obtaining a large set of repeated measures; in the second situation, one studies a large set of people who have been measured once (or a few times). In a measurement model for intra-individual processes, one considers a person-specific measurement model that relates differences in the observed variables (as they occur over time) to a person-specific attribute structure (which varies over time). In a measurement model for inter-individual differences, one considers a model that relates differences in the observed variables (as they occur across people) to an inter-individual differences structure (which describes variation among people).

What does it take for inter- and intra-individual measurement structures to 'converge', in the sense that they arrive on the same conclusions with respect to the measurement model and latent structure? Clearly, this can happen only if the intra-individual differences structure does not differ markedly across persons, for otherwise we need person-specific measurement theories. In addition, it would be beneficial if the intra-individual measurement model and the inter-individual measurement model were isomorphic, so that the measurement model for, say, extraversion, would also obtain within each individual. Hamaker, Molenaar, and Dolan (2005) call this condition *homology*. In that case, for instance, one could say that extraversion is a 'human universal' in the strong sense that everybody's behavior (insofar as it relevant to extraversion) is a function of the same latent structure, much like everybody's length measurements are a function of the same latent structure (i.e., bodily height).

It is sometimes thought that this inference is automatic, so that there is no problem here. The idea underlying this assumption is that evidence for a given factor structure, as derived from inter-individual differences data, is *by itself* evidence for an isomorphic structure 'in the head' of the individual people that make up the population. Examples of this line of thinking are Krueger (1999), who thinks that factors defined in an inter-individual differences context represent 'core psychological processes' that underlie various mental disorders; Kanazawa (2004), who thinks that evidence for general intelligence (the *g*-factor) is also evidence for an adaptation in the form of a single 'psychological mechanism' designed by evolution to solve a particular type of problems; and McCrae and Costa (2008, p. 288), who think that evidence for the Big Five, as derived from the inter-individual differences in personality test data, is also evidence for intra-individual statements like 'E[xtraversion] causes party-going in individuals'.

Such inferences, however, make sense only if there is a logical connection between hypotheses that concern intra-individual and inter-individual levels; i.e., it requires the kind of theoretical system that Cronbach (1957) imagined and Lykken (1991) doubted. In the past 10 years, the idea that such a connection exists as a matter of logical necessity has been refuted by Molenaar and his colleagues. In short, Molenaar and others have conducted simulation studies aimed at showing that standard factor analyses of variation in populations are insensitive to within-subject heterogeneity.

For instance, Molenaar, Huizenga, and Nesselroade (2003) simulate N persons, each of whose behavior is specified by a different factor structure (up to 4 factors). One person may obey a 1-factor structure, another a 2-factor structure, and each person is associated with different factor loadings and error covariance matrix. Thus, with respect to within-subject variability, there is radical heterogeneity. The question, then, is whether there is a between-subject factor model that adequately describes the between-subject variability. If so, then local homogeneity is violated because not every member of the population could exemplify the between-subject model. Molenaar found that a 1-factor structure was sufficient to fit the between-subject variability.

This is, at first sight, surprising because most subjects' time-series data were (by construction) not fit by a 1-factor model and for those whose behavior was specified by a 1-factor model, the factor loadings and measurement-error variances of the between-subject analysis did not match those associated with the time-series data. On a more thorough analysis, however, it is clear that such results may arise, because the between-subjects covariance matrix is partly a function of differences in mean levels of subjects on the observed variables (e.g., this is a variant of Simpson's paradox as displayed in Fig. 4.1; see also Hamaker et al., 2007; Muthén, 1989). In another simulation, Molenaar (1999) determined the factor scores for each subject on the basis of the between-subject model and correlated those scores with the factor scores derived from the time-series data. The correlations were low and in some cases negative. This is also a variant of Simpson's paradox; if the majority of the people with a high mean level on the observed scores are, at a given time point, mostly below their personal means, the relevant correlations become negative.

These simulations show that even the most impressive fit of a between-subjects model to inter-individual differences data does not have implications for the structure of psychological attributes or processes that operate at the level of the individual. Theories concerning that structure are therefore grossly underdetermined by evidence taken from the structure of inter-individual differences. In general, the converse also holds.

Many psychometricians and psychologists, for instance, would guess that if everybody *did* have the same factor structure governing the time series development, then we should find that structure in the inter-individual differences data. That is, if everybody's data come from a person-specific single factor model, then we should find that factor model in the inter-individual differences analysis. Even this, however, is not generally the case. Hamaker et al. (2007) show that arbitrarily complex between-subjects structures can be generated by appropriate manipulations of the averages (over time) around which the time series revolve.

Thus, there is no simple inference ticket from inter-individual differences to intra-individual processes, just as the converse inference ticket does not exist. The accuracy of intra-individual claims on the basis of inter-individual differences research depends on a issue not commonly addressed: whether the measurement models used in the data analysis apply both to differences between people and to differences within people, i.e., are these measurement models homologous?

The conditions for homology to hold are strict. First, it requires local homogeneity, that is, the measurement structure that describes test score covariation for the individual over time must invariant over people. In item response theory, this issue has been addressed by Ellis and Van den Wollenberg (1993), who show that local homogeneity is not implied by standard measurement models for inter-individual differences. In the context of factor analysis, Molenaar, Huizenga, and Nesselroade (2003; see also Molenaar, 1999) have shown the same conclusion to hold.

Local homogeneity refers to the invariance of measurement structures over individuals. Even if such invariance holds, this does not automatically guarantee that the results of an intra-individual analysis will resemble those of an inter-individual differences analysis. That is, if every individual person is adequately described by, say, a single factor model, then one may still find a very different model when analyzing inter-individual differences (Hamaker et al., 2007). The reason for this is that intra-individual time series analyses usually apply to deviations from a person-specific mean, but the covariance matrix of inter-individual differences data is a function of differences between person-specific means as well. The structure of the latter differences is not necessarily constrained by the intra-individual model. Thus, in order to have homology between the inter-individual differences structure, and the results from intra-individual analyses, one needs further conditions to obtain.

First, it appears that to have convergence of the time series structure and the inter-individual differences structure in terms of the dimensionality of the model and the measurement parameters (e.g., factor loadings), one needs not only invariant factor models (which apply to the covariance structure of the data) but also that the data exhibit strict measurement invariance across individuals (which concerns the mean structure; Borsboom & Dolan, 2007; see also Meredith, 1993; Muthén, 1989). This requires that differences in observed mean levels between individuals are exclusively due to differences in latent means. If this is so, then Simpson's paradox cannot occur as it does in Fig. 4.1 or in Hamaker et al. (2007).

We conjecture that these conditions will lead to the same values of the measurement parameters in the measurement model (e.g., factor loadings and error variances in the context of factor analysis), whether it is considered over individuals or over time (Borsboom & Dolan, 2007; Meredith, 1993; Muthén, 1989). However, it need not lead to equivalent values of parameters that describe the latent structure (e.g., means and (co-)variances of latent variables; Muthén, 1989). For a full convergence of the model structures (i.e., including parameters that describe latent variables and relations between them) further conditions are required beyond local homogeneity and measurement invariance. In this case, one needs a condition known as ergodicity (Molenaar, 2004): that is, the results of the analysis as n (the number of persons) approaches infinity must be the same as the results of the analysis as t (the number of time points) approaches infinity. This in turn requires two subconditions. The first condition is *stationarity*: each member of the population ('ensemble') must have stable statistical characteristics, such as a constant mean levels. The second condition is *homogeneity of the ensemble*. If the ensemble is homogeneous, the trajectories of each individual fall under the same dynamical laws. Thus, in this case individuals are fully exchangeable.

As Van Rijn (2008) notes, this is extremely unlikely to describe any situation where inter-individual differences research makes sense. It would imply, for instance, that if 20% of the people have an IQ-score over 115, then every single individual should obtain a score over 115 for 20% of the time. This is clearly nonsensical. In fact, ergodicity cannot hold in cases where stable inter-individual differences exist. This means that whenever there are stable inter-individual differences, the model that describes them will not in its entirety apply to individual. Also, ergodicity will be violated for developmental processes, since they by definition have statistical characteristics that vary over time (e.g., person-specific mean levels are not constant over time). We take this to imply that ergodicity should be viewed as an esoteric condition, that is, we should normally work from the hypothesis that ergodicity does not hold.

The pattern of results that emerges is the following. If ergodicity is violated, but local homogeneity and measurement invariance over individuals hold, then one would expect the dimensionality and measurement model to generalize to the individual, but not the parameters that refer to the latent variables in the model (e.g., means and (co-)variances). If measurement invariance does not hold either, then in addition neither the dimensionality nor the parameters of the measurement model will ordinarily generalize to the individual, although it is still conceptually possible that they will do so by accident (this is a remote possibility). If ergodicity, measurement invariance, and local homogeneity are all violated, then it is impossible in principle for any of the model results to apply at the level of the individual, because the measurement models at the level of the individual and of the population do not match. In this case there is a full disconnect between the proper description of the person and of the population.

The Substantive View: Processes and Inter-Individual Differences

The methodological studies discussed above show that that person-specific measurement models need not be invariant when a between-subjects factor analysis yields a clear pattern. Thus, the various replications of the Big Five personality factors yield some evidence for a between-subject structure, but that evidence is consistent with virtually any hypothesis on person-specific dynamics. It is important to note that the above conclusion concerns the strength of the evidence for person-specific structures as derived from the analysis of inter-individual differences (the strength of this evidence is nil), but that this does not rule out the possibility that ergodicity, measurement invariance, or local homogeneity obtain as a matter of empirical fact. Rather it shows that this is a hypothesis that can only be tested on a case by case basis, by carrying out the relevant research; however, we think that positive results are not to be expected in such research. This becomes clear when one stops to consider the subject matter for areas where these issues are relevant. We will now turn to a discussion of the situation as it obtains in two such areas, namely the study of intelligence and of personality.

The Case of Intelligence

There is no shortage of competing theories of intelligence, but all mainstream theories—and even some of those outside the mainstream such Howard Gardner's (1993) theory of multiple intelligences—posit mental ability (or "intelligence") as a property of individuals. Also, we say things like "John did so well on the test because he's so intelligent" or "Look at how well little Jaime did on her math test; she's so intelligent." Of course, these folk psychological claims are typically completely divorced from substantive psychological theory, but nevertheless, they indicate a commitment to intelligence as some causally efficacious property of individuals. Moreover, these folk psychological claims are not that different from what one finds in a clinical report of one's performance on an IQ test. Therefore, intelligence is plausibly construed as a psychological attribute that applies to the individual. However, psychometric theories of mental ability are based exclusively on between-subject analyses of test performance. They have focused on (differences in) intelligence as a source of inter-individual differences, i.e., differences in intelligence are posited to explain differential performance on tests of mental ability. The obvious and well-worn way to get to the individual from the population is *via* the assumption of local homogeneity, otherwise the tests may be measuring different traits in individuals than in the population. However, given the noted problems in generalizing population structure to the individual, intelligence dimensions like the *g*-factor *cannot* be understood on the basis of between-subject data as denoting mental ability *qua* within-subject attribute.

Psychological practice seems to indicate that psychologists do assume local homogeneity, if only tacitly. The concept of intelligence on which the most popular intelligence tests are based has general intelligence as a central theoretical posit, and general intelligence has its provenance in standard factor analysis of population-level data, not time series analyses of within-subject variability. The commitments of psychometricians are difficult to discern. Famously, Spearman hypothesized that *g* was mental energy, a within-subject attribute. However, he also cautioned his readers that the *g*-factor was only a statistical construct expressing between-subject variability. Jensen, too, does not seem consistent enough to attribute to him a commitment to local homogeneity. Consider the following quote from Jensen (1998, p. 95):

> It is important to understand that *g* is *not* a mental or cognitive process or one of the operating principles of the mind, such as perception, learning, or memory. Every kind of cognitive performance depends upon the operation of some integrated set of processes in the brain. These can be called cognitive processes, information processes, or neural processes. Presumably their operation involves many complex design features of the brain and its neural processes. But these features are not what *g* (or any other psychometric factor) is about. Rather, *g* only reflects some part of the *inter-individual differences* in mental abilities…that undoubtedly depend on the operation of neural processes in the brain.

However in a series of interviews with Frank Miele (2002, pp. 58–59) on the *g-factor* and intelligence, Jensen refers to an individual's *g* as being causally relevant to determining that person future occupational success. Mike Anderson (1992, p. 2) indicates that he assumes local homogeneity when he writes that

> [s]ince differences in tests scores are the target of explanation, whether these represent differences between 2 adults or longitudinal changes within the same individual seems irrelevant. It is taken to be a parsimonious assumption that these differences in scores are to be explained with reference to the *same mechanism*. Thus, for example, higher synaptic efficiency makes on individual more intelligent than another, and increasing synaptic efficiency with age makes us more intelligent as we develop.

Kanazawa (2004) also assumes local homogeneity when he hypothesizes that *g* is a species-typical information processing mechanism (see Borsboom & Dolan, 2006, for a criticism of this position). As indicated, for the *g*-factor to generalize from the population to the individual, local homogeneity is a minimum requirement.

Strictly taken, the model formulation in factor analysis, as it is applied to intelligence data, is not in keeping with the idea of local homogeneity. The problem here is that attributes like general intelligence as supposed to be relatively stable. More precisely, the assumption is that there is little (in practice) or no (in the formulation of standard measurement models) variation in scores across repeated measures for an individual; that is, the latent variable position is usually taken to be a constant for each individual. Typically, variation between testing occasions is attributed to measurement error, not variation in ability.

Psychological theory and psychometric data are often taken to imply that mental ability is stable in this sense, but if it is, then there is no within-subject variability to model, i.e., no time series analysis is available for the individual. With no variability, there is no factor to be extracted. That is, if the standard measurement model were true for intelligence data, such that deviations from person-specific means were solely due to error, then one would expect the analysis of time series data to yield a covariance matrix where all the off-diagonal elements equal zero.

At the population-level, however, we find that the *g-factor* models are robust. As Jensen says in the quoted passage above, "*g* only reflects some part of the *inter-individual differences* in mental abilities". Jensen (2002) makes a more careful statement relevant to the issue of local homogeneity in the context of intelligence research and psychometric models of inter-individual differences:

> It is important to keep in mind the distinction between intelligence and *g*…. The psychology of intelligence could, at least in theory, be based on the study of one person, just as Ebbinghaus discovered some of the laws of learning and memory with N = 1…. Intelligence is an open-ended category for all those mental processes we view as cognitive, such as stimulus apprehension, perception, attention, discrimination, generalization, learning and learning-set acquisition, short-term and long-term memory, inference, thinking, relational education, inductive and deductive reasoning, insight, problem solving, and language. The *g-factor* is something else. *It could never have been discovered with* N = 1, because it reflects inter-individual differences in performance on tests or tasks that involve any one or more of the processes just referred to as intelligence (pp. 40–41, italics added).

That is, *g* is a between-subject statistic, and what it purportedly denotes is a between-subject attribute that "explains" the positive manifold (also a between-subjects phenomenon). The fact of heterogeneity, however, does not imply that the between-subject source of variability is not also a source of variability within subjects. Consider the attribute *height*. Height seems to be an attribute that explains both within-subject and between-subject variability on certain measures such as being able to ride a roller coaster, retrieving items from high shelves, and shoe size. With general intelligence, however, all we have are between-subject models which tell us nothing about how the attribute functions in individuals. Therefore, to make inferences about individual's "general intelligence" being a causal factor is, arguably, unwarranted. Individuals may have some attribute that we can identify as indicative of "intelligence", but the between-subject model does not tell us if it is the attribute purportedly indicated by the *g*-factor. Even though those within-subject attributes may be related to general intelligence, this relationship is not implied by the model.

Apart from the evidence from between-subjects analyses, are there substantive reasons that would lead us to suspect any relevance of an attribute like general intelligence at the level of the individual? Hardly. There is fairly robust evidence that human cognitive development is characterized by stagewise transitions, for instance, which are inconsistent with an interpretation of *g* as a person-specific attribute, because they involve categorical, qualitative steps in development rather than children moving up along a smooth continuum (Jansen & Van der Maas, 2002). Similarly, analyses of various cognitive tasks suggest that mastery of qualitatively distinct rules is needed to solve, say, Raven items, which may also be viewed as a problem for the idea that performance on such tasks is determined by smooth continuum (Verguts & De Boeck, 2002). Language development may likewise be characterized by sudden jumps in understanding (Van Geert, 1991), for instance when children start mastering grammar. In addition, although various reductionist ideas have been put forward, there is no robust evidence for any simple continuous biological substrate that could fill the gap that a dimension like general intelligence leaves at the level of the individual. In fact, the only dynamic theory that has been proposed to explain the occurrence of the positive manifold of intelligence test scores (Van der Maas et al., 2006), which forms the main evidence for *g*, is based on reciprocal relations between various distinct cognitive processes and does not even contain general intelligence in its description of the data-generating process. In conclusion, there is no substantive evidence that general intelligence describes anything more than a structure of inter-individual differences; and substantive theories on human development are virtually uniformly in contradiction with the idea that cognitive development coud be described as a smooth transition along a unidimensional attribute.

The Case of Personality

If one wants a concrete case of our general point—that psychology's research paradigms continue to divide along experimental/correlational lines—there is no better place to look than the psychology of personality. Decades after Cronbach, the seemingly singular professional field continues to harbor two disciplines (Cervone, 1991, 2004).

Even the reader who does not track developments in this field can easily grasp the nature of this divide, and its implications, through a simple thought experiment. First, think of a personality variable. Next, think of a personality theorist. Then compare the two. The personality variable you thought of likely is along the lines of extraversion, or neuroticism, or something related such as sociability, shyness, or friendliness. The theorist likely is Freud or some 20th-century thinker who was significantly influenced by Freud's work. "Extraversion" and "Freud" are prototypic responses.

Now compare them. The "personality variables" refer to average tendencies in thought and action—to what a person does typically. They usually are called *dispositional* variables because they reference a general inclination, or disposition, to act in a certain manner. By contrast, the personality theory of Freud did not even target, as a phenomenon worthy of investigation, average-level behavioral tendencies. Freud saw variability in action rather than average tendencies as revealing of personality. In psychoanalysis one would not, for example, average together "hostility toward same-sex parent" and "hostility toward opposite-sex parent" to gauge a persons "average hostile tendencies." Furthermore, Freud recognized that people engage in superficially similar actions for different underlying reasons; sometimes reasons are related complexly and symbolically to overt emotion and action, and sometimes "it's just a cigar." Average behavioral tendencies, then, are an unsure guide to personality structure.

If you had confined your thought-experiment answers to contemporary personality science (Cervone & Mischel, 2002), the divide would still be apparent. Contemporary theorists of course abandon much of the theoretical and meta-theoretical language of psychoanalysis. Yet, like Freud, many target variability in action that is apparent when one observes individuals across social context (Mischel & Shoda, 1995) and recognize that superficially similar dispositional tendencies may reflect different underlying causes (Cervone, 2004). Overt personality characteristics are seen to result from interactions among psychological systems with different functional properties (Kuhl & Koole, 2004). Nonetheless, others continue to posit that "personality structure" is best described by a system of global dispositional variables (e.g., Ashton & Lee, 2007). In these latter approaches, the core unit of analysis refers neither to behavior-in-context nor to underlying psychological systems, dynamics, or functions. The core variables merely describe what people do on average.

How is one to explain these differences? On the one hand, they are closely related to questions of methodology. Investigators who posit global trait variables tend to employ methods that are correlational in nature. Variables generally are identified via factor analysis of inter-individual differences. Those who adopt other perspectives favor other methods, such as case studies (e.g., Freud, 1900; Hermans, 2001) or experiments (e.g., Greenberg, Koole, & Pyszinski, 2004). So methodological choices may drive the differences between theoretical views.

Yet we suspect that methodological choices *sustain* differences rather than being their origin. Theoretical camps professionalize in such a way that a given method is sanctioned, findings that employ the method are publishable when reviewed by the professional in-group, and the body of published findings sustains the theoretical approach, including the careers of those who espouse it. This sociology of science, however, fails to explain how theoretical differences arose in the first place. How can it be that some investigators view global behavioral tendencies as the structure of personality, whereas others explore personality dynamics and view idiosyncratic, contextualized patterns of variability in action as the key markers of underlying personality structure? It would appear that the very meaning of "personality" and "personality structure" differs from one group of investigators to another (Cervone, 2005). In one case, the terms reference the architecture of mental systems that contribute to those aspects of experience and action that conventionally are called “personality”; this meaning has been apparent since the work of Freud (1923) and remains evident today (Cervone, et al., 2008). In the other, personality constructs serve as a “descriptive taxonomy” (John & Srivastava, 1999, p. 103), and the entity being described is variation in the population at large. How could such divergent conceptions of “personality” have arisen in the first place?

Another thought experiment may be informative. For simplicity, we will shift our focus from persons to an artifact whose properties are fully understood. Suppose that two teams of extraterrestrial investigators landed on Earth and explored what might appear to be dominant large species roaming the land: automobiles. Suppose that one team examined individual automobiles in detail, perhaps with each member of the team taking a close look at a couple of cars, examined one-at-a-time. After this data gathering, members of the group might compare notes to develop a conceptual model of cars. If the extraterrestrials have a good head on their shoulders, they might surmise from their observations that cars have a number of distinct functional systems: a system for storing fuel; a system for burning the fuel; a cooling system; a transmission system; etc. Now imagine that the other group, seeking to save some time, decides to observe the entire population of cars (or a large and presumably representative subpopulation) all at once. Here, differences among cars become apparent: they vary in color and shape; some carry a lot of people and others have just two seats; some cars break down whereas others keep running; all of them seem to travel at about the same speed when they’re on the same roadway, but in very particular circumstances some cars seem a lot faster than others; most of them seem to provide a comfortable space for people to set, but some have extra amenities like leather seats and high-quality stereos. When these investigators sit down to summarize what they have learned, they might conclude that words like “sportiness,” “reliability,” and “luxuriousness” summarize differences among the cars.

What happens when the two research teams meet up? Do the results “converge”; does one “integrate” them? This clearly depends on what the words “converge” and “integrate” are taken to mean. The results do not “diverge.” They are not inconsistent with one another, and they are related in some ways. If one were to pick a between-automobiles dimension such as “sportiness,” and then were to examine mechanical features of those cars that were particularly high and low on that dimension, the cars would differ mechanically. The sporty cars, for example, might have more cylinders and thus generate more power via the burning of fuel. They might also have fewer seats. Yet the two sets of findings do not come together at one conceptual point; they do not combine into a whole (typical meanings of “converge” and “integrate”). They have only a loose association. Terms like “sporty” and “luxurious” are very useful for the purpose of discussing differences among cars. But they do not figure in a conceptual model of what a car has, mechanically, and how the car works.

This analogy maps quite closely to both the history and the current conceptual status of alternative approaches to personality psychology. Historically, some theorists observed individual people in great detail. Freud (1900) conducted case studies. Social learning theorists observed individual children as they acquired skills via interaction with the social environment (Bandura & Walters, 1963). These close observations led them, when providing conceptual models of the person, to model structures, processes, and functions of the human mind. It commonly went without saying for these investigators that a model of "personality structure" was a model of the cognitive and affective systems possessed by the individual (Mischel & Shoda, 1995). At a functional level, they modeled human capabilities (Bandura, 1986).

Other researchers investigate large populations, with each research participant studied only at one point in time and in little depth. Perhaps the best known example of such work is the "lexical tradition" in personality psychology (Ashton & Lee, 2007; Goldberg, 1993). Investigators ask large numbers of persons to describe themselves using personality terms that one finds in the dictionary. Factor analysis is then used to identify dimensions that summarize inter-individual variation. For these investigators, it goes without saying that "personality" refers to differences between people, and "personality structure" is a set of dimensions that summarizes between-person differences in the population at large.

Many efforts in contemporary personality psychology claim to "integrate" these two perspectives. Yet, with the risks of painting with a broad brushstroke, it can be said that these efforts commonly are integrative only in the way that the study of "sportiness" and auto mechanics is integrated in our example above. There is no one-to-one mapping from one language to the other. Innumerable research findings in personality psychology document that people with different scores on between-person trait dimensions differ from one another when those persons are brought into the laboratory and their cognitive or physiological responses to stimuli are assessed (e.g., Eysenck, 1970). Yet, similarly, one could bring cars high and low on "sportiness" into the shop to have their mechanical workings assessed and find that the cars differ. There is only very limited sense in which such findings would "integrate" the two types of research on cars—or persons. And this is not a shortcoming of the research. They can't be integrated into one converging whole. As Harré (1998) has explained with particular clarity, a psychological model of the individual needs to identify the personal powers through which persons think and act. Descriptive terms ("outgoing," "anxiety-prone," "conscientious," and the like) are necessary to social discourse about persons, but one should be very careful in using such terms as cited causes in the explanation of the actions of the individual.

The Conceptual View: Is a Unified Psychology Possible?

The case of personality psychology, then, illustrates the more general point we stated earlier. Many investigators in the field write as if between-person correlational findings have direct meaning for the psychology of individual. In some cases, this intellectual move from inter-individual correlational findings to intra-individual hypotheses is explicit (e.g., McCrae & Costa, 1996, 2008). In numerous other cases, it is a bit more subtle. Researchers may search for the psychological dynamics—i.e., a conceptual model of the individual—that is associated with the given score on a personality trait factor—where the factor summarizes intra-individual differences. For example, they may seek to uncover the psychological dynamics of "introverts" and "extraverts," that is, people with low and high scores on an extraversion scale. This search is sensible if one can assume that the different people who get the same test score are psychologically homogenous. As we saw earlier, there commonly are no grounds for making this assumption.

It is clear from the discussion so far that the gulf that exists between research on intra-individual processes versus research on inter-individual differences is more than a matter of different methodological inclinations, or of researchers' lack of attendance to the project of unification. There appear to be rather principled problems in connecting results from both areas of study. These problems become apparent if one stops to consider the relevant measurement structures in both fields. It is clear that these need not have anything in common. In addition, substantive theories on, say, the dynamics of behavior do not match or support theories on inter-individual differences in behavior; likewise, theories on the development of cognition have no place for such a thing as general intelligence. It is interesting to note, in this respect, that theories of inter-individual differences are not in any relevant sense *refuted* by these observations. In contrast, theory and research on intra-individual processes appears to be largely *irrelevant* to the study of inter-individual differences, and vice versa. The reason is that, barring perhaps the most basic laboratory tasks for which assumptions like ergodicity or measurement invariance over individuals might be taken to hold true, *any* theory on intra-individual processes is compatible with *any* theory of inter-individual differences.

Many people find this to be perplexing. Obviously, the item responses on which inter-individual differences researchers execute their analyses are necessarily generated by some dynamic process in the individual. Also, it is evident that some of the inter-individual differences that researchers find are extremely robust. Furthermore, any set of inter-individual differences is parasitic on the dynamic processes that generated the basic behavior that people exhibit. If John shows up at every other party, while Jane never leaves the house, then clearly there is a dynamic process that differs between them: John does not mysteriously appear at a party without some antecedent dynamic process that, obviously, Jane does not follow. Similarly, if Jane can solve a polynomial equation while John cannot, there must be a process that she carries out but he does not. So how could we have stable inter-individual differences if there were no systematic differences in whatever dynamics describe the actions of the individual?

We think that the answer to this question may be that, instead of there being *no* connection between these levels of analysis, there may actually be *too many*. To see that this may be the case, note that all that is required for a between-subjects measurement model to hold is that (a) there be some set of differences between them that is accurately described by the latent structure, and (b) these differences connect to the observables in the right way, which means that differences in the attribute structure systematically lead to differences in the observables.

Thus, for the hypothesis of general intelligence to be true in the context of the factor model, what is required is that people can be ordered on a line, and that where they are on the line determines their probability distribution over the item responses in the way the model says it does. The model has nothing to say, however, on (a) why or how people come to occupy different positions on this line, or (b) how they produce the answers to IQ-items. That is, John and Jane may have an equal standing on the latent structure called the *g*-factor, but for different reasons. Jane may, for instance, have a smaller brain volume but compensate by having a higher level of neural plasticity, to name but two biological substrates that have been suggested for the *g*-factor (Garlick, 2002; Posthuma et al., 2002). Similarly, both may have a higher probability of answering Raven items correctly than, say, Pete, who has a small brain with low plasticity; nevertheless, they may follow different strategies in answering these items, shaped by different previous experiences and maturation processes. In fact, it is entirely possible that Pete follows the same strategy as John, but is less efficient in his use of memory resources, so that he fails an item where John succeeds. Jane, on the other hand, may follow a strategy different from both John and Pete, and succeed. As long as the processes in play do not affect different items differently (or do so to a sufficiently small degree), there is nothing in the above situation that would falsify a measurement model for inter-individual differences, for the simple reason that such a model makes no claims with respect to the substantive nature of the latent variables it posits or the relations they bear to the observations. It only says that *if* differences arise (in whatever way), *then* these differences must affect the items people take in keeping with the model structure. And this can often happen in an infinity of ways.

It is useful to illustrate how this may work by returning to the automobile metaphor used in the previous section, and exploring it in some more detail. Consider a set of vehicles—say, cars, bicycles, and horse carriages. We may attach to these vehicles an abstract latent structure that refers to a dispositional attribute that determines their performance in races—we call this ‘power’ or ‘maximum performance’, or ‘racing ability’. We may measure this latent structure, for instance by letting the vehicles race on various tracks, using the times needed to complete the tracks as indicators. It is easy to imagine a set of tracks that would show positive intercorrelations analogous to those observed on intelligence test scores: on average, vehicles that perform better on one track will also perform better on other tracks. It is also reasonable to interpret racing ability as a dimension that is real, in the sense that, say, a Ferrari F60 really does have a higher racing ability than a horse carriage with respect to a given set of race tracks (naturally, this does not apply to small mountain paths). One may furthermore suppose that these differences determine differences between the vehicles’ performance, so that the race performances are valid measures of racing ability.

However, if a researcher should set out to determine what ‘racing ability’ consists of, or where it is ‘located’ in the cars and horses under consideration, she would find nothing. Similarly, research into the processes that give rise to differences in performance would probably reveal a bewildering complexity of findings, as these processes differ across vehicles in a myriad of ways. And, should the researcher set out to investigate which physical determinants ‘underlie’ differences in racing ability, the project would strand hopelessly, because the different vehicles have little—in anything—in common when it comes to the propulsion mechanisms that realize their racing ability.

The interesting thing is that all this would not happen because there is *no* relation between the physical processes involved in propulsion and the dispositional attribute of racing ability (there obviously are such relations), but because these relations are themselves dependent on the object under study. The relations involved do not possess sufficient systematicity, generality, and are too complex to allow for a parsimonious explanation of differences in racing ability in terms of the processes that underlie it. Thus, even though there must, by necessity, be processes that underlie differences in racing ability, models that describe inter-individual differences in racing ability and models that describe mechanisms of propulsion for any given vehicle would cover surprisingly little common ground. Moreover, it is very hard to see a way in which a theory on the propulsion mechanism of individual vehicles would place significant restrictions on the model structure that applies to the measurement of racing ability as an interinter-individual differences dimension. In fact, one could imagine that any set of propulsion mechanisms, or of time series models describing them, would be consistent with any structure of inter-individual differences.

It is thus likely, should there be car scientists that consider such questions, that they should develop intra-individual and inter-individual research traditions as psychologists have. And it is questionable, as in the case of psychology, whether the intra-individual and inter-individual twains would ever meet. To us, the situation sketched in the car example thus appears to be quite similar to the situation as it exists in the fields that show the greatest tension between intra-individual and inter-individual levels of analysis, such as personality and intelligence research. General intelligence, for instance, is extremely similar to racing ability. Personality traits like extraversion are similar as well, although they are not maximum performance concepts but typical performance concepts; thus, such traits would bear more similarity to notions such as ‘reliability’, as explained in the previous paragraph.

Why are Inter-Individual Differences Intractable?

The question that arises is: what properties of such inter-individual attributes lead them to separate themselves so clearly from the intra-individual analysis? We think that three properties are important in this respect: their dispositional character, the fact that they are multiply realizable, and the fact that they are multiply determined.

First, almost all inter-individual differences concepts are essentially dispositional. That is, their meaning relies heavily on an ‘if…then…’ structure. The typical example of a dispositional concept, for instance, is ‘fragility’. To say that a vase is fragile is to say that it has a physical structure that leads it to break if it is dropped. Whatever physical structure precisely realizes the property of fragility is not relevant to the truth-value of the sentence ‘this vase is fragile’. For intelligence, such ‘if…then…’ relations are filled in like ‘John is highly intelligent: if he is presented with a difficult problem, he will solve it’. For personality traits, they are filled in like ‘John is extraverted: if he were given the choice between staying at home with a book or going to a party, he would choose the latter’. It does not matter for the truth-value of such conditionals precisely *how* John solves items or gets to parties. Also, it does not matter *what* allows or forces him to exhibit such behaviors. In fact, these concepts are amenable to a functionalist analysis, in the sense that it may be upheld that, at the level of the individual, *whatever* allows him or her to solve an item in an IQ test *is* intelligence. Thus, in this sense concepts like intelligence, extraversion, and racing ability are essentially open; that is, they can be (physically) realized in infinitely many ways.

This points to a second important property of inter-individual differences dimensions, which is that their levels can be often expected to be multiply realizable. Just like a given level of racing ability can be realized by different vehicles in different ways, a given level of intelligence may be realized in different people in different ways. To see this, it is illustrative to note that, should we tomorrow be visited by little green men from outer space who, instead of a brain, have a hydraulic system located in their left big toe that does the thinking, they might still be located on the dimension of general intelligence as long as their levels of intelligence can be placed on the same line as ours and behave in the same way, even though the item response processes, at a physical level, may have few elements in common with our own. This thought experiment, naturally, represents an extreme case, but it is in our view highly likely that in the human population general intelligence (if it exists) is realized differently in different people as well; this appears to be almost guaranteed by the sheer complexity of the human brain and the existence of inter-individual differences in cognitive and emotional development. Such different realizations of the levels of inter-individual differences dimensions can be expected to involve ‘physical’ differences (e.g., in the context of intelligence, brain size, neaural plasticity, neural connectivity, etc.) as well as ‘psychological’ ones (e.g., differences in strategy, the use of cognitive rules and heuristics, etc.).

A related but distinct property of inter-individual differences dimensions is that they are not just multiply realizable (the same level of intelligence may be realized by different constitutions) but also multiply determined: the causal pathways that lead to any given level of an inter-individual differences dimension are likely to differ among people. There is ample reason to expect this to be so. For instance, the combination of (a) high heritability estimates for almost all inter-individual differences dimensions (Boomsma, Busjahn, & Peltonen, 2002) and (b) the limited success in finding any genetic markers that explain more than, say, 1.5% of the variance in such dimensions, suggests that inter-individual differences may be strongly polygenic. This is evidence for multiple determination as far as it concerns the part of development that is under genetic control, because it means that distinct pathways underlie inter-individual differences for (almost) any distinct combination of individuals. Another source of evidence for multiple determination comes from the study of epigenetic effects (Jaenisch & Bird, 2003; Molenaar, Boomsma, & Dolan, 1993), which is an autonomously operating process that creates inter-individual differences that are not uniformly tractable to any set of genes or environmental conditions. Finally, at the environmental side of development, the differential pathways that lead to equivalent levels of ability are completely obvious. To give an example, John and Jane may have the same level of intelligence at a given time point, because Jane may have had a virus of accident that impaired her intelligence to equal the initially lower level of John, whose intelligence has undergone no major impairments. Any such external influences, insofar as they do not distort the measurement model for a given test, must be counted as part of the causes that give rise to the inter-individual differences dimensions under study; and it is clear that their number is infinite. Taken together, the evidence suggests that our working assumption should be that inter-individual differences stand under the influence of a large number of disparate causal factors.

We think that it is plausible to assume that most inter-individual differences variables are dispositional, multiply realizable, and multiply determined. The implication of this is that, even though each and every difference between two people depends for its existence on *some* differences in intra-individual processes, the systematic explication of the relation between these domains is likely to be an extremely complicated matter; in fact, in many cases, this relation may be intractable. This observation is consistent with the psychometric analysis discussed earlier in this chapter, which established the lack of correspondence between inter-individual differences structures and the structure of intra-individual processes. Thus, although causally dependent on intra-individual processes, inter-individual differences may not lend themselves to an explanation in terms of these intra-individual processes. This, in our view, may be one of the reasons that the two disciplines of scientific psychology, as discussed by Cronbach 1957, have not appreciably moved closer. In fact, we suspect that the character of the relation between intraindividual processes and inter-individual differences may serve to isolate these branches of study from each other in a structural way.

Supervenience

The reason for this is that the relation between intraindividual differences and inter-individual processes, as explicated in this chapter, is most aptly characterized as a *supervenience* relation. A property X supervenes on a (set of) properties Y if and only if it is true that, given a fixed Y, there cannot be differences in X. A typical supervenience relation in psychometrics, for instance, is that of the relation of a total score (X) to the item scores (Y) of which it is composed: there cannot be differences in the total score if there are no differences in the item scores. The supervenience relation is asymmetric, as can be easily seen from the same example: if there are no differences in the value of the total score (X), there may nevertheless be differences in the item scores (Y). This is because the total scores are multiply realizable, as for n items, a total score k can be realized in $n!/\{k!(n-k)!\}$ ways.

Together with multiple realizability, the supervenience relation has been used often in the literature on the mind-body problem to give a nonreductive physicalist account of the relation between mental states and brain states. Roughly, physicalism holds that all mental phenomena *are* ultimately physical phenomena. Reductionism holds that, in addition, psychological laws and regularities can ultimately be *reduced* to (or systematically explained in terms of) physical theories, for instance to those concerning the human brain. Thus, physicalism is an ontological thesis and reductionism is an epistemological one. Nonreductive physicalism roughly holds that psychological states (like, for instance, 'believing that π is not a rational number') can be realized in an infinite number of ways in the human brain. Thus, although there cannot be differences in psychological states if there are no differences in the physical structures that realize them (supervenience and materialism), there may be differences in the physical structures that serve to realize the same psychological state (multiple realizability). The primary argument against reductionism that follows from this (explicated by Fodor, 1974) is that the physical category of states that realize a psychological state will be arbitrary from the perspective of the reducing theory (say, neuroscience) and therefore cannot figure it its laws.

We submit that the relation between intra- and inter-individual differences is exactly the same as that between mental and physical processes. That is, every inter-individual difference depends, for its existence, on a difference in intra-individual processes (supervenience). However, these differences are multiply realizable, which means that the intra-individual processes that 'realize' a given level of intelligence only do so from the perspective of the higher level science (inter-individual differences research). They do not form a homogeneous category from the perspective of the lower-level science (intra-individual processes). Therefore, the collection of intra-individual processes that is contained in the correlational psychologist's 'has intelligence level *x*' is not a consistent category from the perspective of the experimental psychologist: from the perspective of the experimental psychologist, it corresponds to a disjunctive 'either follows process *a*, or *b*, or *c*, or…', and this disjunction is arbitrary from an intra-individual processes perspective. Therefore, it will not be a 'kind' of intra-individual research, and cannot figure in its laws.

Illustration: The Case of Chess Expertise

The related issues of multiple realizability, multiple determination and the dispositional character of intra-individual cognitive abilities are present in a wide range of psychologically interesting concepts. An almost archetypical example of a cognitive process, playing chess, illustrates how these three elements interact to make intra-individual inferences from interindividual data improbable, if not impossible.

Chess playing is a psychologically interesting skill that encompasses a variety of cognitive skills and processes, much in the same way as IQ can be seen as combination of skills that yields an individual score with predictive qualities. The equivalent of chess IQ is the international rating system called the Elo-rating, after the

American physicist Arpad Elo. Although the distribution of scores is logistic rather than normal, the overall nature of the Elo-rating is very similar to the IQ score. An individual has a score that is a rank on a unidimensional ability scale, which reflects the probability of beating lower or higher ranked individuals, and the likelihood of solving chess problems of varying complexity. A closer examination will show that all three previously discussed issues hold for chess playing.

First, chess playing and chess ability are essentially dispositional. In principle, there are no limits to the cognitive process, playing style or set of abilities a player uses to win games; all that matters is the ratio of wins and losses against variably skilled opponents and the probability of solving problems. Players of comparable chess playing ability may constitute their respective levels in very different manners; one player may possess a vast knowledge of common situations and by-the-book tactics, whereas another may rely more on intuition and creativity. As long as they have the same scores on the Elo-scale, there is nothing on the inter-individual level to set them apart, which allows for rather dissimilar processes to fall under the umbrella of ‘chess playing at level x’.

In addition, evidence from the neurosciences suggests that chess ability is a multiply realized ability, even on the intra-individual level (over time). An example is a study by Amidzic, Riehle, Fehr, Wienbruch, and Elbert (2001), in which magneto-encephalogram recordings (MEG) were made of both expert chess players and intermediate players whilst playing a chess computer. The patterns of cortical activity for 5 s after the computer made a move were recorded and compared. Amateur chess players showed pronounced temporal lobe activity, a region commonly associated with logical reasoning skills such as ‘if… then…’ statements. The pattern for experts (ELO>2000) was markedly different. They showed very little temporal activity but pronounced prefrontal lobe activity, which is normally related to memory and retrieval activity while intermediate players showed mainly temporal lobe activity. This result was very robust, and showed a strong negative correlation (–0.84) between Elo-rating and activity in medial temporal lobes, the perirhinal and entorhinal cortex and related structures. It is known that expert chess players are able to memorize the patterns that often occur in chess matches up to a staggering 100,000 and 400,000 moves or situations (De Groot, 1978). This suggests that as a player becomes better, he or she relies more and more on ‘pre-programmed’ positions, so that deciding on the next best moves becomes much more a memory activity than a reasoning ability. This is a prime example of a cognitive ability that shows significant qualitative changes not captured by the interindividual model. It seems therefore that chess playing ability is a multiply realizable skill; there are many ways to play chess and they change markedly with increased skill. Finally, chess playing is multiply determined. There is a wide range of skills that are useful when playing chess, but the interplay between them is potentially very complex and not suitable for simple factor analytic approaches. For example, an increase in working memory capacity may only be an advantage if one’s knowledge of strategy allows for the efficient use of this extra capacity.

It seems clear that the causal factors that contribute to the overall quality of a chess player are irreducible on several different levels. It must be stressed that this is not an exotic exception to the rule, if anything, chess is exemplary for a wide variety of cognitive abilities that psychologists deem worthy of study. These issues do not preclude a coherent analysis, but awareness of measurement issues are an essential safeguard against overly ambitious intra-individual inferences drawn from any form of group level measurement.

Clearly, the dispositional character of inter-individual differences dimensions, together with multiple realizability and multiple determination, yields significant problems for attempts to sensibly connect these dimensions to intra-individual processes. This appears to grant such dimensions a certain sense of autonomy and irreducibility. For instance, it has been argued in the literature that multiple realizability is a sufficient condition to block successful reduction of the higher-level theory to the lower-level theory; Fodor (1974, 1997) famously maintains that this holds for higher-level sciences as diverse as psychology, economics, and meteorology. This conclusion has been hotly debated in the philosophical literature of the past three decades, and it is beyond the scope of this chapter to evaluate its validity. However, apart from the principled question whether reduction is at all possible, we think it is relatively obvious that the existence of supervenience and multiple realizability will seriously complicate the practical integration of fields.

Conclusion

It has been the working assumption of many psychologists and methodologists that the integration of experimental and correlational research or, if you will, intra-individual processes and inter-individual differences research, is a matter of time; that it is a sign of the 'immaturity' of psychology that they have not yet converged to a single theoretical system; and that the unification of psychology is something that we should strive for. The image that arises from the present investigation, however, is a rather different one. The rift separating the traditions may be much deeper than is commonly thought and, in fact, may be structural—that is, the gap will not be closed by the passing of time or the progression of scientific psychology. It may very well be here to stay. Thus, to speak with Fodor (1974), we may want to accept not the unity, but the disunity of psychology as a working hypothesis.

The evidence for this hypothesis is quite overwhelming. First, the fact is that more than 50 years have passed since Cronbach's call for integration, and that they have done so without widespread progress being made in this particular program. Naturally, one may consider various explanations of this situation that draw on sociological processes (e.g., the formation of research traditions) or differences in methodological orientation (as Cronbach himself did by labeling the traditions as 'correlational' and 'experimental'). However, we seriously doubt whether such explanations have sufficient explanatory force. Scientists tend to relentlessly pursue lines of research that 'work', in the sense that they answer interesting questions or lead to the solution of practical problems, and it seems rather implausible that so few 'working' versions of the desired integration had been stumbled upon if they were there for the taking. The traditions of 'correlational' and 'experimental' research may not be induced by different methodological inclinations, but by a different subject matter.

Moreover, psychometric considerations suggest that few restrictions on one side of the divide can be deduced from theories that apply to the other side: a particular dimension of inter-individual differences can be generated by many systems of intra-individual processes, and conversely a theory of intra-individual processes does not lead to restrictions on the possible spaces of inter-individual differences unless unreasonably strong restrictions are met. For instance, Hamaker et al. (2007) and Timmerman et al. (2009—this book) show how far little intra-individual and inter-individual structures can diverge. The only restriction that is universally in place is that intra-individual and inter-individual theories should be *consistent* with each other—in the sense of not being contradictory—and the psychometric work of the past few decades strongly suggests that this restriction is extremely easy to meet. However, mere consistency of theories is far to little to fuel an integration of fields, or to drive an explanation of inter-individual differences in terms of intra-individual processes. Psychology is entirely consistent with, say, non-Euclidean geometry, but that does not imply that there are any interesting explanatory connections between these areas of research.

To have a real connection between the fields under consideration here, one should be able to infer what an inter-individual differences structure should like from a theory of intra-individual processes—more specifically, one should be able to place refutable restrictions on the inter-individual model structure. This is certainly not impossible in general, but for many sub-disciplines in psychology the task at hand appears to be extremely difficult to carry out. More specifically, the sort of attributes that inter-individual differences research has brought into play appear to be of the wrong kind to figure in such explanatory schemes. One may of course counter that this just means that the inter-individual differences attributes should be done away with, and replaced by process-oriented theories. This, however, requires one to actually show that such replacements will work adequately, and this need not be possible. Returning to the intelligence example, for instance, there have been several proposals to fill the gap of things like *g* by substituting sets of cognitive processes at the level of the individual (e.g., Sternberg, 1985), but the empirical success of such approaches has been limited (Deary, 2000) and it is not clear that such process theories are at all in the same explanatory league as inter-individual differences dimensions, in the sense that they may not apply to the same phenomena (e.g., the positive manifold; Borsboom, Mellenbergh, & Van Heerden, 2003). The similarity to the mind-body debate is quite strong in this case as well; for instance, we find similar calls for 'brain-based' constructs instead of 'psychological' ones among the fiercest reductionists (e.g., Churchland, 1981). Such calls, however, are promises; and a general law that applies to promises is that the proof of the pudding is in the eating. Clearly, so far there has been little pudding to eat.

Scientific progress comes in many forms. The textbook example is the successful explanation of a phenomenon in terms of a theory, but sometimes science progresses by showing that a dreamed route of progress is blocked. Famous examples include

Gödel's (1931) incompleteness theorem, which destroyed the work presented in Russell and Whitehead's (1910) *Principia Mathematica* by showing that the desired reduction of mathematics to logic was impossible, and the theory of complex systems, which for instance explains why we cannot predict the weather more than a few days in advance. Our suggestion in the present work is that the integration of intra-individual and inter-individual research programs may be exactly such a case: a dreamed route of progress that is really a dead end street.

This may sound like a gloomy conclusion. However, we think that there is little reason for optimism on the 'integration' of the two disciplines of psychology in the sense Cronbach (1957) had in mind, and wishful thinking is not bound to change that. Moreover, there are two important implications that follow from the analysis, if it is correct, that may serve to further our understanding of how the disciplines could be related. The first implication is that we need further understanding on the conceptual and empirical relationships between attributes as they are used in the two disciplines. We have established, reasonably firmly, that equating the concepts of intra-individual processes research and inter-individual differences research is not an option that we should expect to work. At the same time, it would seem that the experimental psychologists 'working memory' and the differential psychologists 'working memory' are related, and how they may be is a important issue. Clearly, we have only scratched the surface with respect to this interesting question. Second, the present analysis cautions against interpreting results from inter-individual differences research as descriptive of the individual person; similar caution should go out to most experimental studies, which are descriptive of means, not individuals. Thus, the analysis of the individual in its own right is a project that, despite a century of psychology, still awaits a proper methodology. It is our hope that methodological techniques suitable to this purpose will be developed to maturity in the coming years.

Acknowledgments Denny Borsboom's work was supported by NWO innovational research grant no. 452-07-005.

References

Amidzic, O., Riehle, H. J., Fehr, T., Wienbruch, C., & Elbert, T. (2001). Pattern of focal γ-bursts in chess players. *Nature 412*, 603.

Anderson, M. (1992). *Intelligence and development*. Cambridge: Blackwell.

Ashton, M. C., & Lee, K. (2007). The HEXACO model of personality structure and the importance of the H factor. *Social and Personality Psychology Compass*, *2*, 1952–1962.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., & Walters, R. H. (1963). *Social learning and personality development.* New York: Holt, Rinehart, & Winston.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244.

Boomsma, D. I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics*, *3*, 872–882.
Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.
Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*, 1089–1108.
Borsboom, D., & Dolan, C. V. (2006). Why g is not an adaptation: A comment on Kanazawa. *Psychological Review*, *113*, 433–437.
Borsboom, D., & Dolan, C. V. (2007). Theoretical equivalence, measurement, invariance, and the idiographic filter. *Measurement*, *5*, 236–263.
Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219.
Cervone, D. (1991). The two disciplines of personality psychology. *Psychological Science*, *2*, 371–377.
Cervone, D. (2004). The architecture of personality. *Psychological Review, 111*, 183–204.
Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, *56*, 423–452.
Cervone, D., Caldwell, T. L., Fiori, M., Orom, H., Shadel, W. G., & Kassel, J., et al. (2008). What underlies appraisals?: Experimentally testing a knowledge-and-appraisal model of personality architecture among smokers contemplating high-risk situations. *Journal of Personality*, *76*, 929–967.
Cervone, D., & Mischel, W. (Eds.). (2002). *Advances in personality science*. New York: Guilford Press.
Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, *78*, 67–90.
Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 612–684.
Danziger, K. (2000). Making social psychology experimental: A conceptual history, 1920–1970. *Journal of the History of the Behavioral Sciences*, *36*, 329–347.
Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford: Oxford University Press.
De Groot, A. D. (1978). *Thought and choice in chess*. Den Haag, Mouton.
Ellis, J. L., & Wollenberg, A. L. v. d. (1993). Local homogeneity in latent trait models: A characterization of the homogeneous monotone IRT model. *Psychometrika*, *58*, 417–429.
Eysenck, H. J. (1970). *The structure of personality*. (3rd ed.). London: Methuen.
Fodor, J. (1974). The special sciences, or: The disunity of psychology as a working hypothesis. *Synthese*, *28*, 97–115.
Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Nous*, *31*, 149–163.
Freud, S. (1900/1953). *The interpretation of dreams. Standard Edition* (Vol. 4, pp. 1–338; Vol. 5, pp. 339–621). London: Hogarth Press.
Freud, S. (1923/1961). *The ego and the id. Standard Edition* (Vol. 18, pp. 12–66). London: Hogarth Press.
Gardner, H. (1993). *Frames of mind*. New York: Basic books.
Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of inter-individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*, 116–136.
Greenberg, J., Koole, S. L., & Pyszczynski, T. (Eds.). (2004). *Handbook of experimental existential psychology*. New York: Guilford Press.
Gödel, K. (1931). Über formal unentscheidbare Satze der Principia Mathematica und verwandter Systeme I. In Feferman, S. (1986). *Kurt Gödel: Collected Works*, (Vol. 1, pp. 144–195). Oxford: Oxford University Press.
Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*, 26–34.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate time series analysis. *Multivariate Behavior Research*, *40*, 207–233.

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait-state model. *Journal of Research in Personality*, *41*, 295–315.

Harré, R. (1998). *The singular self: An introduction to the psychology of personhood*. London: Sage.

Hermans, H. J. M. (2001). The construction of a personal position repertoire: Method and practice. *Culture and Psychology*, *7*, 323–365.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, *35*, 245–254.

Jansen, B. R. J., & Van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*, 383–416.

Jensen, A. (1998). *The g factor*. Westport: Praeger.

Jensen, A. (2002). Psychometric g: Definition and substantiation. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 39–54). Mahwah: Erlbaum.

John, O. P., & Srivastava, S. (1999). The big-five factor taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford Press.

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, *111*, 512–523.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry*, *56*, 921–926.

Kuhl, J., & Koole, S. L. (2004). Workings of the will: A functional approach. In J. Greenberg, S. L. Koole, & T. Pyszczynski (Eds.), *Handbook of experimental existential psychology* (pp. 411–430). New York: Guilford.

Lamiell, J. T. (1987). *The psychology of personality: An epistemological inquiry*. New York: Columbia University Press.

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology. Vol. 1: Matters of public interest* (pp. 3–39). Minneapolis, MN: University of Minnesota Press.

McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51–87). New York: Guilford.

McCrae, R. R., & Costa, P. T., Jr. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Sage handbook of personality theory and assessment* (Vol. 1, pp. 273–294). Los Angeles: Sage.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525–543.

Miele, F. (2002). *Intelligence, race, and genetics: Conversations with Arthur Jensen*. Boulder, CO: Westview Press.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–286.

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

Molenaar, P. C. M. (2004). A manifesto on psychology as ideographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218.

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., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Under-*

standing human development: Dialogues with lifespan psychology (pp. 339–360). Dordrecht: Kluwer Academic Publishers.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, England: Cambridge University Press.

Posthuma, D., de Geus, E. J. C., Baaré, W. F. C., Hulshoff Pol, H. E., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience*, *5*, 83–84.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, *13*, 238–241.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (2001). Unified psychology. *American Psychologist*, *56*, 1069–1079.

Timmerman, M. E., Ceulemans, E., Lichtwarck-Aschoff, A., & Vansteeland, K. (2009). Multilevel simultaneous component analysis for studying intra-individual variability and inter-individual differences. In J. Valsiner & P. C. M. Molenaar (Eds.), *Developmental process methodology in the social and developmental sciences* (pp. 291–318). New York: Springer.

Van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, *98*, 3–53.

Van Rijn, P. (2008). *Categorical time series in psychological measurement. University of Amsterdam*, Unpublished doctoral dissertation.

Verguts, T., & De Boeck, P. (2002). The induction of solution rules in Raven's Progressive Matrices test. *European Journal of Cognitive Psychology*, *14*, 521–547.

Whitehead, A. N., & Russell, B. (1910). *Principia mathematica*. Cambridge: Cambridge University Press.