MULTIVARIATE BEHAVIORAL RESEARCH
2024, VOL. 59, NO. 6, 1290–1309

OPEN ACCESS

Integrating Intra- and Interindividual Phenomena in Psychological Theories

Denny Borsboom[a*] and Jonas Haslbeck[a,b*]

[a]Department of Psychological Methods, University of Amsterdam; [b]Department of Clinical Psychological Science, Maastricht University

ABSTRACT

Psychological science is divided into two distinct methodological traditions. One tradition seeks to understand how people function at the individual level, while the other seeks to understand how people differ from each other. Methodologies that have grown out of these traditions typically rely on different sources of data. While both use statistical models to understand the structure of the data, and these models are often similar, Molenaar (2004) showed that results from one type of analysis rarely transfer to the other, unless unrealistic assumptions hold. This raises the question how we may integrate these approaches. In this paper, we argue that formalized theories can be used to connect intra- and interindividual levels of analysis. This connection is indirect, in the sense that the relationship between theory and data is best understood through the intermediate level of phenomena: robust statistical patterns in empirical data. To illustrate this, we introduce a distinction between intra- and interindividual phenomena, and argue that many psychological theories will have implications for both types of phenomena. Formalization provides us with a methodological tool for investigating what kinds of intra- and interindividual phenomena we should expect to find if the theory under consideration were true.

KEYWORDS

Individual differences; intra-individual processes; ergodicity; formal theory

Introduction

The realm of psychology can be divided in two traditions (Cronbach, 1957). The first tradition originates with the work of experimental and mathematical psychologists such as Wundt, Helmholtz, and Fechner, and is dedicated to understanding processes, structures, and attributes that characterize the functioning of the individual person (Murray, 2020). The second tradition originates with the work of Francis Galton, James McKeen Cattell, and Charles Spearman, and studies the origin and structure of individual differences in psychological attributes (Galton, 1879; J Cattell, 1890; Spearman, 1904). These two streams of thought are also visible in the major traditions in mathematical modeling in psychology, where the Wundt line foreshadows mathematical psychology and the Spearman line psychometrics (Wijsen et al., 2019). In a nontrivial sense, these approaches involve a figure-ground reversal: for a scientist who targets universals, individual differences constitute the background, whereas for the scientist who targets individual differences, universals constitute the background. For example, a cognitive scientist interested in the mechanics of working memory may regard individual differences a nuisance, while for a psychometrician interested in measuring working memory capacity, only deviations from the average are of interest (Borsboom et al., 2009).

The question of how to relate these different traditions has been subject to investigation since Cronbach (1957) put the two disciplines of scientific psychology on the map as comprising two distinct, and often competing, methodological traditions. He contrasted the traditions methodologically, by juxtaposing experimental approaches (with the stereotype of a researcher who executes statistical tests on differences in means, manipulated with experiments) and correlational ones (with the stereotype of a researcher computing correlations between sets of individual differences). Many people intuitively expect the qualitative results from such approaches to converge. However, it turns out that this is not necessarily the case, for instance because such patterns can change, and even reverse, as a function of conditioning on a third variable (Kievit et al., 2013; Wagner, 1982) or because sources of variance are tangled up in ways that are not straightforwardly separated (Hamaker, 2012). To use a standard example, one could erroneously conclude that, if forcing people to type faster leads to more errors (experimental approach), then we should also expect that people who type faster should make more errors (correlational approach). But of course a correlational research design will demonstrate that people who type faster typically make less errors, and that is the case because there is a third variable—typing ability—that produces this correlation: better typists are both faster and more accurate. The methodological understanding of this type of reversal, and of relations between experimental manipulations and correlation structures generally, has been greatly advanced since the advent of modern theories of causality (Holland, 1986; Pearl, 2009; Peters et al., 2017; Rohrer & Murayama, 2023; Weinberger, 2015).

Next to the contrast between intraindividual experimental approaches and interindividual correlational traditions, a similar question comes up when contrasting intraindividual correlations from a single individual to corresponding interindividual correlations. Specifying under which conditions these two types of evidence converge has become a central question since intensive longitudinal (or time series) data have become highly prevalent in psychological research (Conner & Barrett, 2012; Hamaker et al., 2016; Hamaker & Wichers, 2017; Kuppens et al., 2022; Miller, 2012; Trull & Ebner-Priemer, 2014). In a seminal manifesto, Molenaar (2004) argued that a necessary condition for this convergence is ergodicity. This condition implies that time series of different people should both be stationary (implying, for example, that there is no trend in the data) and homogeneous (the behavior of all individuals is governed by the same generating model). In this case, individuals are to a considerable extent exchangeable, much like particles in a gas are exchangeable save for position and momentum. Molenaar's paper raised the important point that, for systems that are non-ergodic, we cannot make a straightforward interindividual → intraindividual inference (and vice versa). This is because, unless the system under study satisfies ergodicity, statistical patterns characterizing the individuals will not be the same as the statistical patterns characterizing interindividual differences. This generalizes to other properties of the probability distributions in question, such as the dimensionality and parameters of latent variable models that describe these distributions (Hamaker, 2012; Hamaker et al., 2007; Molenaar et al., 2003).

Since ergodicity is unlikely to hold in many of the systems studied by psychological research, Molenaar's manifesto showed that characterizing statistical patterns in individuals requires analyzing data of individual subjects. This in turn requires a reasonably large number of observations of the individual at hand, which led to a sharp increase of intensive longitudinal (or time series) analysis in psychological research. This methodological approach has flourished, and has obtained an important place in the arsenal of psychological methods (Bringmann et al., 2017; Cabrieto et al., 2017; de Haan-Rietdijk et al., 2017; Fisher et al., 2017; Gates & Molenaar, 2012; Hamaker, 2012; Haslbeck et al., 2021; Haslbeck & Ryan, 2022; Wichers et al., 2016, 2020, 2015). A benefit of this fast development of such methodologies has been a surge in the development and application of intraindividual analyses to empirical data (e.g., Contreras et al., 2019; Fisher et al., 2018, 2017; Robinaugh et al., 2020; Wichers et al., 2016), which allowed psychological scientists to study the temporal evolution of psychological processes with a granularity that was before typically unfeasible. Methodological understanding of the issue has also increased. It is now well-known that statistically separating individual differences from intraindividual processes requires both multiple time series and dedicated analyses on these time series (Adolf et al., 2014; Hamaker, 2012; Hamaker et al., 2016, 2015; Mulder & Hamaker, 2021). Thus, a direct statistical inference from one level to the other is now generally precluded.

However, the fact that direct statistical inference from one level to the other is generally not possible does not mean that statistical patterns from both levels cannot be integrated. After all, we have learned about many intraindividual causal processes from interindividual statistical patterns; examples include the discoveries that smoking causes cancer and that childhood trauma contributes to adult psychopathology. Evidence for such claims rests largely on the systematic analysis of individual differences, rather than on experimental manipulation and intraindividual time series data. Similarly, we have learned how individual differences can arise from studying intraindividual processes; for example, we now know that large individual differences can be produced purely through extensive practice; one way in which this has been shown is through case studies in which individuals devoted thousands of hours of deliberate practice into developing their skills (Ericsson, 2008). These examples show that, even in cases where ergodicity may be violated, one can find ways of integrate evidence drawn from the intra- and interindividual levels when developing a theory. In this paper, we develop a conceptual framework that systematizes how theories relate to statistical patterns on the intra- and interindividual levels and how theories can be developed using both.

In a nutshell, we will argue that most psychological theories make predictions about both statistical patterns that one would expect to see in time series that characterize the individual, and statistical patterns one would expect to find in individual differences data. Since theories imply statistical patterns at both the intra- and interindividual levels, we can use these statistical patterns to inform the theory. We will argue that this type of inference back to the theory can take the form of constraining the space of theories, which is a weaker form of inference than the direct statistical inference approach. We will also argue that deriving the statistical patterns on different levels that are implied by a theory is all but impossible with verbal theorizing and making predictions using intuition. Instead, we think that formalized psychological theories are needed, because such theories can be used to deduce the precise implications of the theory at both the intra- and interindividual levels through mathematical analysis and simulation. As a result, formalization can be a tool to deduce what one should expect to see in intra- and interindividual comparisons, given the theory. The fact that we can precisely deduce which patterns on the intra- and interindividual level are implied by our theory also means that we can test and develop the theory based on all deduced patterns.

The organization of this paper is as follows. First, we discuss how the traditional distinction between time series and individual differences emanates from two ways of contrasting the individual: with themselves at other time points, and with other people at the same time point. We propose that such contrasts lead to qualitatively different types of phenomena, which we call intraindividual phenomena and interindividual phenomena. We suggest that many psychological theories carry implications for both types of phenomena, even in cases where statistical models that are used to capture these phenomena in the data are neither equivalent nor transparently related. Subsequently, we discuss how to connect theories to intra- and interindividual phenomena. Here, we focus on an indirect form of inference in which phenomena constrain the set of theories that are consistent with them. Using examples from different psychological disciplines, we show that this type of inference—albeit weaker than direct statistical inference—can be used to develop theories with all types of phenomena. We then illustrate the fact that a theory can make predictions about different types of phenomena and therefore can be developed based on these phenomena using a recently proposed formal theory of panic disorder. We conclude by discussing how taking the perspective of theory, phenomena and data allows us to integrate disparate empirical research and thereby strengthen theory development in psychology.

The conceptual architecture of intra- and interindividual phenomena

Many discussions about the relation between intraindividual processes and interindividual differences start with Cattell's data box (R. B. Cattell, 1988). This box crosses individuals, variables, and measurement occasions. Intraindividual comparisons can be made using a slice of the data box in which one individual is studied over multiple occasions; interindividual comparisons can be made using a slice in which multiple individuals are studied at a single occasion. Paradigmatic intraindividual approaches arise from case studies, for instance using a number of repeated administrations of a task or intensive longitudinal (or time series) data obtained through Experience Sampling Methodology (ESM; Hektner et al., 2007). Paradigmatic interindividual approaches involve cross-sectional designs, in which a large number of individuals are assessed at a single time point. These designs can be combined, for instance when a number of individuals are measured at many occasions; a multiple time series design. It has been often shown in the literature that statistical analysis of these different slices can lead to different conclusions (Hamaker, 2012; Molenaar, 2004; Molenaar et al., 2003).

As such, the conceptual space in which the intraindividual vs. interindividual issue is discussed is organized in terms of the relation between statistical models (e.g., the factor model) and data (e.g., cross-sectional data versus time series data). This line of analysis typically concerns models that are statistically identified and estimable, which allows one to show that models estimated on intra- versus interindividual data would yield different results. In cases where models do not align, a central objective of modeling is to separate interindividual and intraindividual sources of variance so as to arrive at parameter estimates that can be clearly interpreted at either level (Hamaker, 2012; Mulder & Hamaker, 2021).

Framing the intra-interindividual distinction in terms of statistical inference is highly useful and helps us avoid unwarranted inferences from one domain to the other. However, applying the same reasoning to other types of inference may hinder research. This becomes clear when focusing on theory development as a core activity of science and acknowledging that many theories are not statistical models. Examples of such theories could be verbal theories or formal theories that are not uniquely identified or estimable by a single type of dataset. Here, work in philosophy of science (Bogen & Woodward, 1988; Woodward, 2011) suggests that theories do not explain a specific dataset, but instead theory explains phenomena evidenced by the data. For instance, Darwin's theory of evolution does not explain any particular observation of a finch's beak, but a robust correlation between beak size and environment; Newton's gravitation laws do now explain the path of one individual apple falling, but the general shape of the trajectories of falling bodies; Spearman's g-factor theory does not explain why John answered an IQ-item correctly, but the general feature that IQ-items tend to be positively correlated. As these examples suggest, the phenomena that serve as explanatory targets for scientific theories often take the form of empirical generalizations (Haig, 2008). In statistics, such empirical generalizations are typically considered in terms of statistical inferences from a sample to a population. This construction will serve us well in the current paper, too. In the remainder of this section, we will introduce the distinction between data, phenomena and theory; define intra- and interindividual phenomena, and relate them to different psychological theories.

The theory-data-phenomena distinction

Data

In statistical modeling, the term *data* is commonly used to indicate a dataset, as for instance stored in a spreadsheet; Cattell's data box is a good example of such a setup. An entry in a dataset contains a representation of an observation that was made. For instance, if the data point corresponding to the i-th row and the j-th column contains the symbol "1", that symbol may represent the observation that person i gave an answer to item j that we evaluated as "correct". Two features of data are important for our current purposes. First, that data are particular (Haig, 2008). That is, any data point represents an observation that was made by some person at some place at some time. Second, that an entry in a datafile has representational content. This means that it is not itself an observation, but rather a symbolic representation of an observation. In the natural sciences, one can often glance over this issue and treat the data as if they were direct observations (a tradition also common in statistics). However, in psychology, this is not a wise course of action, because responses often depend on how research participants understand the questions that are included in a questionnaire. This is relevant in the present context, because questionnaire items may themselves contain references to temporal dynamics; for instance, when a symptom questionnaire item queries the presence of a symptom over the past months (Ryan and Dablander, n.d.). This is important for the relation between the data and the processes we study, because explicit and implicit time references in the items should be taken into account when interpreting statistical relations in the data.

Phenomena

By structuring the representations of observations in systematic ways, researchers create variables. The construction of variables involves an additional step of abstraction, in which relations between the symbolic representations in the data are used to facilitate the study of patterns in the data. The simplest of these relations is equivalence. For instance, we would typically consider all entries of the symbol "1" in a column of testing data to represent that the persons in question answered the item correctly (even though their actual responses may have been different, as in open question formats). This results in equivalence classes that make up the simplest type of variable (usually called "nominal"). More complicated variables can be constructed by using additional relations between equivalence classes, as in the case of ordinal, interval, or ratio scaled variables (Krantz et al., 1971). Statistics is largely concerned with modeling the joint distribution of sets of variables that are constructed in this manner.

In the current paper, we will take phenomena to be robust patterns that characterize these distributions. For instance, a univariate example would be the bimodality in children's responses to conservation tasks (Van der Maas & Molenaar, 1992); a multivariate example would be the pattern of positive correlations between cognitive test scores that is known as the positive manifold (Van der Maas et al., 2006). The most important feature of phenomena, for our current purposes, is that in contrast to data they are *general* (Haig, 2008). For example, the bimodality in conservation task responses is a general feature that is evidenced in data patterns obtained from different populations in different locations at different time points. Importantly, that phenomena invariably involve generalization does not mean that statistical patterns must be invariant over all conceivable domains of generalization; in fact, a violation of invariance can constitute an important phenomenon in itself. For instance, robust differences between cohorts, such as the secular gains in IQ known as the Flynn effect (Trahan et al., 2014), constitute a robust phenomenon in the study of intelligence, and cross-cultural differences may similary consitute robust phenomena in and of themselves (Henrich et al., 2010).

Because phenomena are generalizations, they typically cannot be established with certainty on the basis a particular data set. As such, the relation between data and phenomena should be construed as *evidential*: data provide evidence for phenomena (Borsboom et al., 2021; Haig, 2008). As a result, the tasks of detecting phenomena and determining to what extent they generalize across subpopulations, cultures, and time points is an important part of the scientific enterprise.

Theory

Scientific theories can be understood in many different ways. In the current paper, we work with a relatively elementary understanding in which theories are interpreted more or less literally. This means we take scientific theories as attempts to characterize the structure of the world we study. Thus, when Spearman says that individual differences in test scores arise from individual differences in mental energy, we take him to mean that there is actually such a thing as mental energy, and the amount of energy one has plays a decisive role in causally generating the correct responses to items on an intelligence test.

This literal understanding of theories is known as scientific realism (Devitt, 2005), and it is typically contrasted with alternative understandings in which theories are variously interpreted as predictive instruments or "inference tickets" (Ryle, 1949), devices that allow us to navigate the world pragmatically (James, 1909), or highly efficient ways of representing data, as in logical positivism (Suppe, 1977). The advantage of a realist understanding of theories in the present context is that it allows us to reason under the assumption that the theory is true in a relatively straightforward manner. Namely, we can build an artificial "world" in which the theory is true, and study what phenomena we would expect in that scenario. This aligns with the common use of simulation in statistical modeling, where modelers investigate what we should expect to see in the data if a given model were true. We can use a very similar setup to investigate what phenomena (patterns in the data) would follow, if the theory were true, namely by simulating a world in which that is the case. If that simulation exhibits the empirical phenomena we want to explain, then the theory putatively explains the phenomena (van Dongen et al., 2022). The explanation is putative because the premises (the theoretical model simulated from) may be incorrect; moving from a putative to a correct explanation involves additional research that supports the premises of the argument.

Thus, while the relation between data and phenomena is evidential (data provide evidence for phenomena), the relation between theory and phenomena is explanatory (theories explain phenomena). The relation between theory and data is thus mediated by phenomena, in the sense that a theory draws support from the data indirectly, namely by offering a good explanation for the phenomena the data evidence (Haig, 2008).

Intra- versus interindividual phenomena

The conceptual architecture provided by the theory-phenomena-data distinction sheds a different light on how to integrate data patterns from intra- and interindividual contrasts. This is because of the intermediate level of phenomena. If theories are conceptualized as statistical models that directly provide a likelihood associated with each data point, then the natural way of relating theories and data is through statistical estimation. From this perspective, the focus provided by Molenaar (2004) is instructive, because it shows that *if* theories are considered to be identifiable statistical models, *then* the adequate estimation of intraindividual parameters from interindividual or combined data either requires strong statistical conditions to hold (Adolf et al., 2014; Molenaar, 2004), or specialized applications of multilevel models to separate parameters that characterize the individual's time series from parameters that characterize differences between individuals (Hamaker, 2012; Mulder & Hamaker, 2021).

Intra- and interindividual contrasts

However, if explanatory theories are understood as targeting phenomena that are evidenced by the data, the situation is different. From this point of view, the way that an individual's current responses are contrasted with other responses defines different variables. For example, the variable constructed by taking a slice from Cattel's data box in the time direction contrasts the individual's current responses with that individual's responses at different time points (e.g., as in an ESM study). However, if we construct a variable differently, by taking a slice in the individual differences direction, we contrast the individual's responses with other individuals' responses (e.g., as in a standard $T = 1$, $N =$ large, psychometric study). These are not the same variables, because the representational content of the data being represented is not the same. That is, the data in the different slices of the data box have different meanings, and therefore the variables constructed from them have different meanings too. In this respect, it is actually confusing that often the same term is used to indicate both, for instance when a model for individual differences is compared to one for intraindividual time series, and variables in both models are indicated by the same word, e.g. 'depressed mood' (Bos et al., 2017).

Now, because the variables defined in this way are different, the statistical patterns that they form can be different as well. That is, they provide evidence for qualitatively different phenomena. Generally, we would expect the statistical patterns of those phenomena to be different. However, they might be the same in very specific cases. One of those is the case in which statistical patterns captured by a uniquely identified statistical model and in which ergodicity holds. For this reason, we propose that it is useful to distinguish between intra- and interindividual phenomena, where intraindividual phenomena are statistical patterns that characterize an individual, and interindividual phenomena are patterns that characterize individual differences (see Figure 1):

Importantly, in many cases, data-analytic procedures followed to identify empirical phenomena will involve both contrasts in some way. We denote such empirical phenomena by the term *combined phenomena*. Such phenomena include relations between individual differences in person-specific means of time series and other individual differences, as when person-specific means in a time series of depression symptoms are related to neuroticism or gender.

The distinction between the different types of phenomena is best illustrated through examples. One paradigmatic example of an intraindividual phenomenon is the speed-accuracy tradeoff (Heitz, 2014). The speed accuracy tradeoff refers to the phenomenon that, if tasks are executed under increased time pressure, the probability of errors in the task increases. This is an intraindividual phenomenon, because it denotes a statistical pattern (a negative correlation between speed and accuracy) that is formed by contrasting the individual's responses to other responses by that same individual. Thus, an intraindividual phenomenon can, in principle, be demonstrated as a statistical pattern in data gathered in a single person.

A paradigmatic example of an interindividual phenomenon is the positive manifold of intelligence (Jensen, 1999; Spearman, 1904; Van der Maas et al., 2006). The positive manifold is a statistical pattern of robust positive correlations between performance on different cognitive tasks. These correlations are computed over variables that represent responses of different individuals; thus, this is a contrast between the individual's responses and other individuals' responses. Hence, this is an interindividual phenomenon. An interindividual phenomenon cannot be demonstrated in the data of a single individual, because it requires the comparison of the individual with other individuals.

Combined phenomena, which involve both intraindivdiual and interindividual contrasts, commonly arise from statistical approaches that simultaneously use both intra- and interindividual contrasts; as a consequence, many phenomena that are established through, say, multilevel modeling of multiple time series may be viewed as combined phenomena; this includes phenomena that involve interindividual contrasts that are corrected for intraindividual processes, as would for instance arise from applications of the random intercept cross-lagged panel model (Hamaker et al., 2015).

A paradigmatic example of a combined phenomenon is the predictive relation between autoregressive coefficients for affect states in ESM time series and depression diagnoses (Kuppens et al., 2010; van de Leemput et al., 2014). Here, an intraindividual phenomenon is first identified in the form of the correlation between affect states at subsequent time points, where the individual's current response is contrasted with other responses by the same individual. Then, in a second step, the differences between these autoregressive coefficients are conjoined with depression diagnoses to form a higher order statistical pattern: a positive correlation between the value of these coefficients and depression diagnoses.

In the following section we will use intra- and interindividual phenomena and discuss how they can be related to theories at the intra- and interindividual levels.

Connecting theories to intra- and interindividual phenomena

In order to connect to intra- and interindividual phenomena, a theory needs to have implications for these

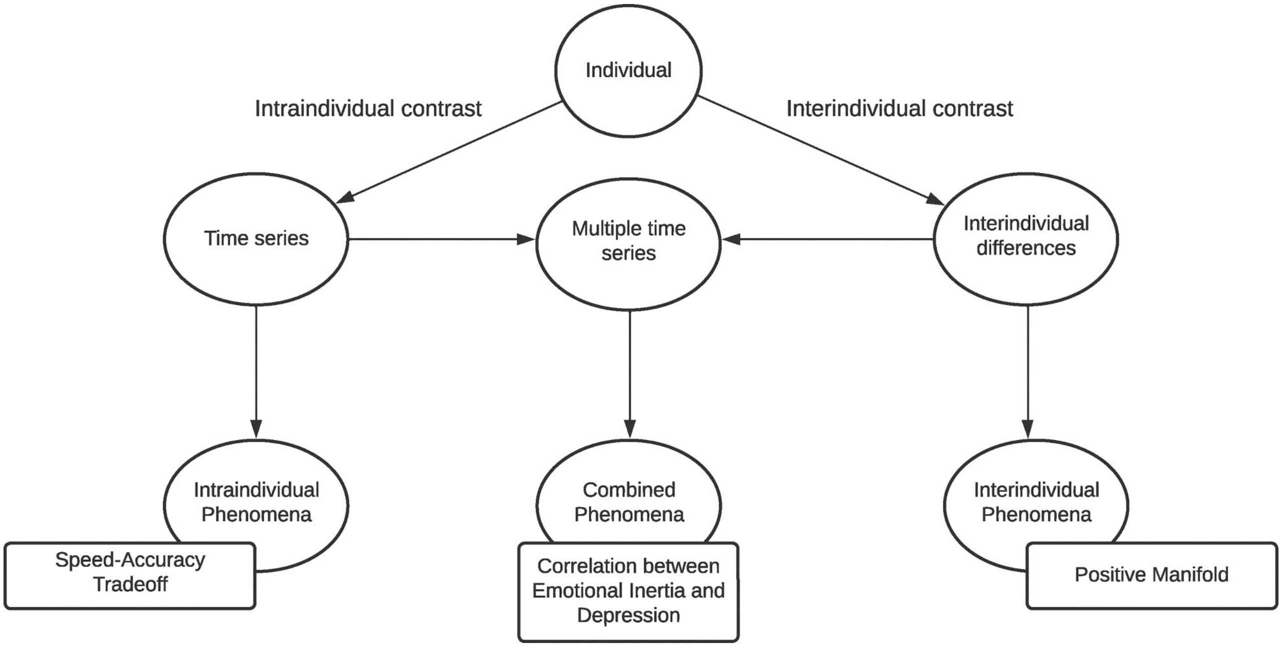

Figure 1. Three kinds of phenomena in psychological research. Intraindividual phenomena are statistical patterns that result from contrasting the individual's current state with other time points. Interindividual phenomena result from contrasting the individual with other individuals. Combined phenomena are statistical patterns that result from a combination of both contrasts. Prototypical examples are listed in boxes.

phenomena. To achieve this, theories about how individuals function (henceforth: intraindividual theories) have to specify a data generating mechanism for the individual person. This specification is akin to the intraindividual tradition of experimental psychology, as described by Cronbach (1957), and to time-series traditions in psychology. Theories about individual differences (henceforth: interindividual theories) need to specify how the data generating mechanism leads to individual differences. This specification is akin to the interindividual tradition of correlational psychology and its psychometric arsenal of methods.

In the present context, a key question is how interindividual theories relate to intraindividual phenomena, and how intraindividual theories relate to interindividual phenomena. As shown by Molenaar et al. (2003), this relation cannot be made through direct statistical inference, since the level of the theory is not aligned with the level of the phenomenon. However, we argue that a relation can be established through a weaker "indirect" form of inference, which takes the form of constraints that phenomena on both levels can put on a given theory. We first discuss this type of indirect inference in more detail. We then cross the intraindividual-interindividual distinction with the theory-phenomena distinction, and evaluating how the relevant phenomena can be be related to intra- and interindividual elements in psychological theories using different types of inferences.

Direct statistical inference vs. indirect types of inferences

In many popular methodological approaches in psychology, such as Item Response Theory (IRT) and Structural Equation Modeling (SEM), a preferred strategy for connecting theories to data is to construct a statistical model that is structurally similar to the theory. For instance, in confirmatory SEM, models are explicitly constructed in such a way that they resemble one's theory as much as possible. Ideally, such a model is uniquely identified and testable against the data using a goodness-of-fit test. This approach is sometimes called one of *direct inference*, because the model makes contact with the data directly, typically using a likelihood function that assigns probabilities to all possible data patterns.

Using direct statistical inference of this kind limits the class of models that can be used to connect theories to data. This is particularly difficult if we aim to test intraindividual theories using interindividual phenomena, as the requisite statistical models require the data to fulfill highly restrictive conditions (Adolf et al., 2014; Hamaker et al., 2007; Molenaar, 2004). In many cases, even though a theory can make predictions about phenomena at different levels, we cannot estimate or test the theory directly on the basis of a single dataset. This may be the case because the theory is unlikely to be isomorphic to *any* statistical model that may be uniquely identified by the data; for instance, because the theory is too complicated to be translated into an identified model.

However, it is still possible to assess the plausibility of a theory through its ability to explain empirical phenomena. This strategy is one of *indirect inference* (Haslbeck et al., 2022; Hosseinichimeh et al., 2016), both because the inferential relation between data and theory is mediated by statistical patterns (i.e., empirical phenomena) and because it does not uniquely identify the theory. The key concept to reason through this setting is the one of constraints. If we can imagine a set of theories S that predict, say, an interindividual phenomenon that we do *not* observe in the data at hand, then this finding constrains the set of all theories such that we can exclude S. The more informative the phenomena are with respect to the theory at hand, the stronger this constraint will be. For example, we would expect that, in general, intraindividual time series data will provide phenomena that result in stronger constraints for a system evolving over time than interindividual data. The complement of this set, which includes the theories whose predictions *are* consistent with a given set of phenomena, could be interpreted as an equivalence class. This concept is well-known to researchers familiar with causal inference, where the (conditional) dependencies in multivariate data constrain the causal graph only to an equivalence class containing all, for example, Directed Acyclic Graphs (DAGs, e.g., Peters et al., 2017) that could have produced the observed dependency structure.

In the remainder of this section, we investigate whether connections between intra- and interindividual theories and intra- and interindividual phenomena can be made using the strategy of indirect inference.

Intraindividual theories and intraindividual phenomena

The relation between intraindividual theories and intraindividual phenomena is the most straightforward one from the perspective of psychology, which tends to focus on processes that characterize the individual human being. It epitomizes the earliest traditions of psychology (Murray, 2020), in which case studies were instrumental, for instance to test theories of functions like perception, sensation, and memory.

If intraindividual phenomena are to be explained, the system that the theory characterizes (e.g., memory) and the phenomena (e.g., forgetting curves) are at the same level. This means that a theory of the target system can be relatively directly aligned with phenomena evidenced by the data. In modern psychology, this strategy is visible in the area of mathematical psychology. A good example concerns research into the drift-diffusion model (Ratcliff & McKoon, 2008; Wagenmakers et al., 2007), which describes the cognitive integration of evidence in two-choice response tasks, such as the lexical decision task (Ratcliff et al., 2004). The model is used to explain intraindividual phenomena, such as the speed-accuracy tradeoff (van der Maas et al., 2011), which are established by analyzing large numbers of repeated trials that are gathered at the individual level.

In the case of the drift-diffusion model, the model is sufficiently simple that its parameters can be estimated directly from a single dataset (Wagenmakers et al., 2007). In other cases, intraindividual theories are too complicated to achieve this; examples of such a theoretical systems are the ACT-R framework (Anderson, 1996) and neuroscientific models like Dynamic Field Theory (Bhat et al., 2022). In these cases, model parameters are tuned by triangulating findings from different research designs, data sets, and other sources of information. Because the models are not directly identifiable, they are typically tested indirectly, *via* their capacity to describe and explain experimentally established phenomena.

Intraindividual theories and interindividual phenomena

In many situations, it is not possible to align intraindividual theories and intraindividual phenomena because the data needed to establish the intraindividual phenomena are unavailable. This may be the case for various reasons; for instance, the intraindividual theory may connect events at a time scale that extends beyond the reach of empirical studies, the theory may concern unique and singular events that only occur once so that no variation should be expected at the intraindividual level, or we may lack assessment techniques that would be required to gather relevant data at the level of the individual.

Still, in many cases where it is impossible to establish intraindividual phenomena, researchers do evaluate intraindividual theories, and they typically do this by comparing different individuals. Perhaps the most salient example of this approach is the randomized experimental design, in which causal effects are estimated by comparing mean differences between experimental conditions. In such a design, counterfactual causal reasoning is used to create an evidential link between intraindividual theory and interindividual phenomena (robust differences in sample means). For example, in causal inference, the population mean of the control group may be assumed to equal the population mean of the experimental group under the counterfactual assumption that the latter group had not received the treatment (Holland, 1986). Note that the average treatment effect can only be used to make inferences about the treatment effect for an individual if we are willing to assume that all individual treatment effects are the same (Lamiell, 1987); critically, however, even without assuming homogeneity in individual causal effects, the average causal effect can still *constrain* the possible sets of individual causal effects.

In cases where experimental manipulations are infeasible, so that the randomized experimental design cannot be implemented, correlational data may be combined with causal assumptions to furnish a basis for causal inference (Holland, 1986; Pearl, 2009; Rohrer & Murayama, 2023; Weinberger, 2015). For example, in psychology, an important interindividual phenomenon concerns the correlation between individual differences in childhood abuse and individual differences in adult psychopathology. This phenomenon may be explained through an intraindividual theory; for instance, childhood abuse may disrupt attachment, which may lead to distrust in other people, which may lead to paranoia (Isvoranu et al., 2017). This is a case where direct statistical estimation is structurally infeasible, because ethical and practical constraints preclude the gathering of data that could inform such a process.

However, the intraindividual theory that childhood abuse causes adult psychopathology may still connect to the interindividual phenomenon *via* an explanatory link: if we assume that some individuals are exposed to abuse, while others are not, and run a simulation according to this scenario, we may find that the implied differences in means follow from the theory. Although this of course does not allow one to conclude that the theory is correct, the interindividual phenomenon does put *some* constraints on the theory, because not all intraindividual theories could have generated the observed interindividual phenomena; therefore, theories that could not have generated these phenomena lose credibility.

Interindividual theories and interindividual phenomena

Typically, interindividual theories specify how interindividual differences in one dimension (e.g., genetic makeup) cause individual differences in another dimension (e.g., IQ-scores). For instance, common phrasings of the theory of general intelligence hold that (a substantial part of) the individual differences in the *g*-factor are caused by individual differences in genetic makeup, while the resulting differences in the *g*-factor themselves are expressed in a wide variety of cognitive tasks, as for instance included in typical intelligence tests (Jensen, 1999).

Interindividual theories will imply interindividual phenomena naturally, and, as is the case for the scenario where intraindividual processes explain intraindividual phenomena, evaluation of the explanatory merits of a theory is relatively straightforward. However there is one important difference. While intraindividual theories need not reference individual differences at all, individual differences research always contains implicit assumptions about intraindividual processes, because some intraindividual process is invariably necessary to connect the theory to the interindividual phenomena. For instance, for genetic makeup to be expressed in behavior, there must be functioning human bodies producing proteins and growing a brain. Thus, the theory must rely on at least a minimal set of assumptions characterizing the individual person. However, such assumptions take the form of background or auxiliary assumptions, in the sense that the theory is agnostic about them. That is, although genetic theories of the *g*-factor require *some* intraindividual process to be operational, they do not commit to a *particular* intraindividual process. For example, it does not matter for the interindividual theory whether genetic effects on IQ are transferred *via* brain volume (Posthuma et al., 2002), *via* neural plasticity (Garlick, 2002), or *via* a myriad of processes (Kievit et al., 2014); at least one such route must exist, but which one is operational is evidentially neutral with respect to the theory (Weinberger, 2015).

Interindividual theories and intraindividual phenomena

In the previous cases, the link from theory to data could be constructed in terms of implications of the theory. Interindividual theories have implications for interindividual phenomena; intraindividual theories can both have implications for intra- and interindividual phenomena. As a result, these phenomena can constrain the set of candidate theories, because theories that fail to explain them lose credibility.

This is not the case for the combination of interindividual theories and intraindividual phenomena, because even though interindividual theory does implicitly call on some intraindividual process, the nature of this process is not usually an explicit part of the theory. As a result, it seems one cannot derive intraindividual phenomena from such a theory, either through simulation or through mathematical analysis. If this conclusion is correct, interindividual theories cannot explain intraindividual phenomena.

Nevertheless, there are examples in which intraindividual phenomena do provide some evidence for interindividual theories because they support intraindividual theories that are known to be consistent with interindividual theories. An example occurs in research that utilizes a link between the IRT model—a model for interindividual differences—and the drift-diffusion model—a model for intraindividual processes (Tuerlinckx & De Boeck, 2005). In this research line, it has been shown that *if* data are generated through a drift-diffusion model, and individuals differ in the parameters of that model, *then* the resulting patterns of correct and incorrect responses will be described by an Item Response Theory (IRT) model (with the IRT parameters a direct function of the diffusion parameters). As a result, intraindividual phenomena (e.g., the speed-accuracy tradeoff) that indicate that the intraindividual process is in fact described by the drift-diffusion model indirectly bolster confidence in the adequacy of the interindividual theory, because we know that if the intraindividual theory is correct, the interindividual theory must also be correct. Thus, even though the intraindividual model functions as an auxiliary relative to the interindividual theory (the IRT model does not imply the diffusion model for the response process), if we know the intraindividual model has a particular structure then we know the interindividual theory must hold (the drift-diffusion model for the response process does imply the IRT model).

Developing an intraindividual theory with phenomena on all levels

Theories, as we interpret them here, stipulate that the world has a certain structure. The theory putatively explains phenomena if they would follow "as a matter of course" (Peirce, 1931) if the world indeed had that structure. This means that we can derive phenomena at different levels from a theory and use those to test and develop the theory. In this section, we demonstrate this by simulating theory-implied intraindividual, interindividual, and combined phenomena from an intraindividual theory of panic disorder (Robinaugh et al., 2019) and discussing how they can be compared to empirical phenomena to test and develop the theory.

Deriving phenomena on different levels requires formalization

To study this relation, we need to evaluate which phenomena follow from a given theory. If the theory is stated in a purely verbal manner, this is difficult, for two reasons. First, verbal theories are in almost all cases imprecise, by which we mean that different formal theories which can make competing predictions are consistent with them (Farrell & Lewandowsky, 2018). And second, even if the theory is precise, if it involves more than a few components that are related in a non-linear way, humans are generally unable to intuit its implications. The field of complexity science illustrates this, as it contains many studies of simple systems that have unexpected and complex behaviors (Mitchell, 2009). For example, the extremely simple logistic map creates phase transitions and deterministic chaos (May, 1976) and a model consisting of three simple local rules is able to create flocking behavior of birds (Reynolds, 1987).

We think that the systems we are studying in psychological research are almost certainly of that nature, which suggests that verbal theories are too imprecise to yield clear implications and therefore formalized theories are needed (Borsboom et al., 2022; Fried, 2020; Haslbeck et al., 2022; Robinaugh et al., 2021; Smaldino, 2017; van Rooij & Baggio, 2021). A theory that is formalized into a mathematical or computational model will generally allow one to derive statistical patterns that should follow from it, either analytically or through simulation, which is particularly useful in cases where intuition fails. Thus, after we have cast the theory in mathematical form, we can investigate which phenomena follow from the theory, for instance by simulating data under the assumption that the theory is true. In principle, such phenomena could either be intraindividual, interindividual, or combined.

A brief overview of a computational model of panic disorder

Panic disorder is a psychiatric syndrome characterized by recurrent panic attacks and persistent patterns of behavioral changes, such as avoidance of situations that may trigger panic attacks (American Psychiatric Association, 2013). The computational model of panic disorder proposed by Robinaugh et al. (2019) suggests that panic disorder is the result of three interlocking feedback loops, shown in Figure 2.

Firstly, there is an reinforcing feedback loop between Arousal and Perceived Threat (Clark, 1986), in which Arousal can trigger Perceived Threat and Perceived Threat can, in turn, lead to higher Arousal. This feedback loop is moderated by Arousal Schema, which is a set of beliefs and associations regarding the threat posed by Arousal. For instance, a person may hold the belief that increased heart rate and palpitation (Arousal) may be signs of a medical condition, which leads the person to interpret them them as potentially dangerous (Perceived Threat). Thus, when Arousal Schema is high, the effect of Arousal on Perceived Threat is increased and it can create a positive feedback loop that causes substantially elevated levels of Arousal and Perceived Threat, a state known as a panic attack. Panic attacks eventually terminate through the operation of homeostatic feedback processes (H), which serve to restore Arousal to a normal level.

Second, Perceived Threat and Escape Behavior (e.g., fleeing the location where a Panic Attack takes place) are linked through a dampening feedback loop, where heightened Perceived Threat causes an individual to act in a way that reduces the perceived consequences of higher Arousal. As this behavior reduces Perceived Threat, it helps keep Arousal and Perceived Threat in balance. The individual's Escape Schema—their beliefs in their ability to cope with Perceived Threat without engaging in Escape Behavior—moderates the strength of this loop, with higher Escape Schema meaning that already low levels of Perceived Threat can trigger Escape Behavior.

Finally, there is a third feedback loop in the system through which individuals can learn how dangerous Arousal is perceived (Arousal Schema) and the amount of Perceived Threat that can be tolerated or managed while refraining from Escape Behavior (Escape Schema). Unlike the fast feedback between Arousal and Perceived Threat and between Perceived Threat and Escape Behavior, which operates on a time scale of minutes, this third feedback loop is slow and operates on a time scale of days to weeks. It is also essential to note that the learning process, and thus the change of Arousal Schema and Escape Schema, relies upon the collective behavior of Arousal, Perceived Threat, and Escape Behavior (which is signified by the parallel arrows originating from the grey box that encompasses all fast-moving components). If Arousal remains close to equilibrium and Perceived Threat remains low, then learning is not possible and Arousal Schema and Escape Schema remain unaltered. However, if Arousal and Perceived Threat become significantly elevated (e.g., if a panic attack occurs) then learning is possible.

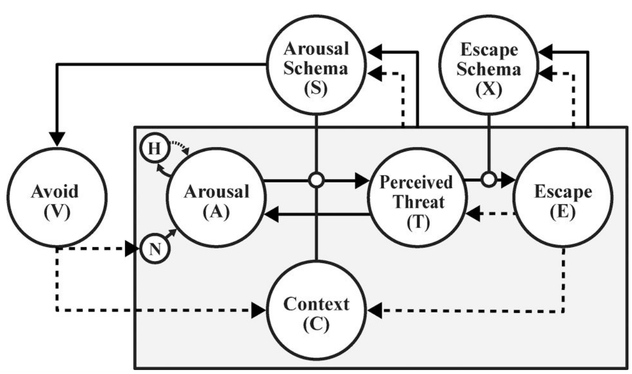

Figure 2. The causal diagram of the computational model of Panic Disorder by Robinaugh et al. (2019). The components within the grey box are fast-changing processes at a time scale of minutes; the components outside are slow-changing processes on a time scale of days.

What is learned critically depends on whether Escape Behavior is shown. If Escape Behavior is not carried out during a panic attack, so that the attack is endured, this may lead the individual to learn that such behavior is not necessary and that Arousal is not something to be feared (increasing Escape Schema and decreasing Arousal Schema). This putatively explains the effectiveness of exposure therapy (Robinaugh et al., 2019). On the other hand, if Escape Behavior is employed, the individual is likely to infer that the anticipated catastrophe would have been realized if they had not taken such action (decreasing Escape Schema and increasing Arousal Schema). This increased Arousal Schema then makes the individual more likely to be exposed to heightened arousal and situations that would lead to further Arousal, while they become more hesitant to enter these situations (*N* and *C* in Figure 2, respectively). Ultimately, this can lead to panic disorder, in which recurrent panic attacks cause a systematic and enduring disruption of a person's life. For a precise description of all components and relationship, and a detailed analysis of the behavior of the model we refer the reader to Robinaugh et al. (2019).

Generating data from the computational model

The fact that the panic disorder model is a computational model implies that all relationships are precisely specified. Given a set of initial values for all variables, this allows us to compute the behavior of the model arbitrarily far into the future. To illustrate how this computational model can imply intraindividual, interindividual and combined phenomena, we simulate

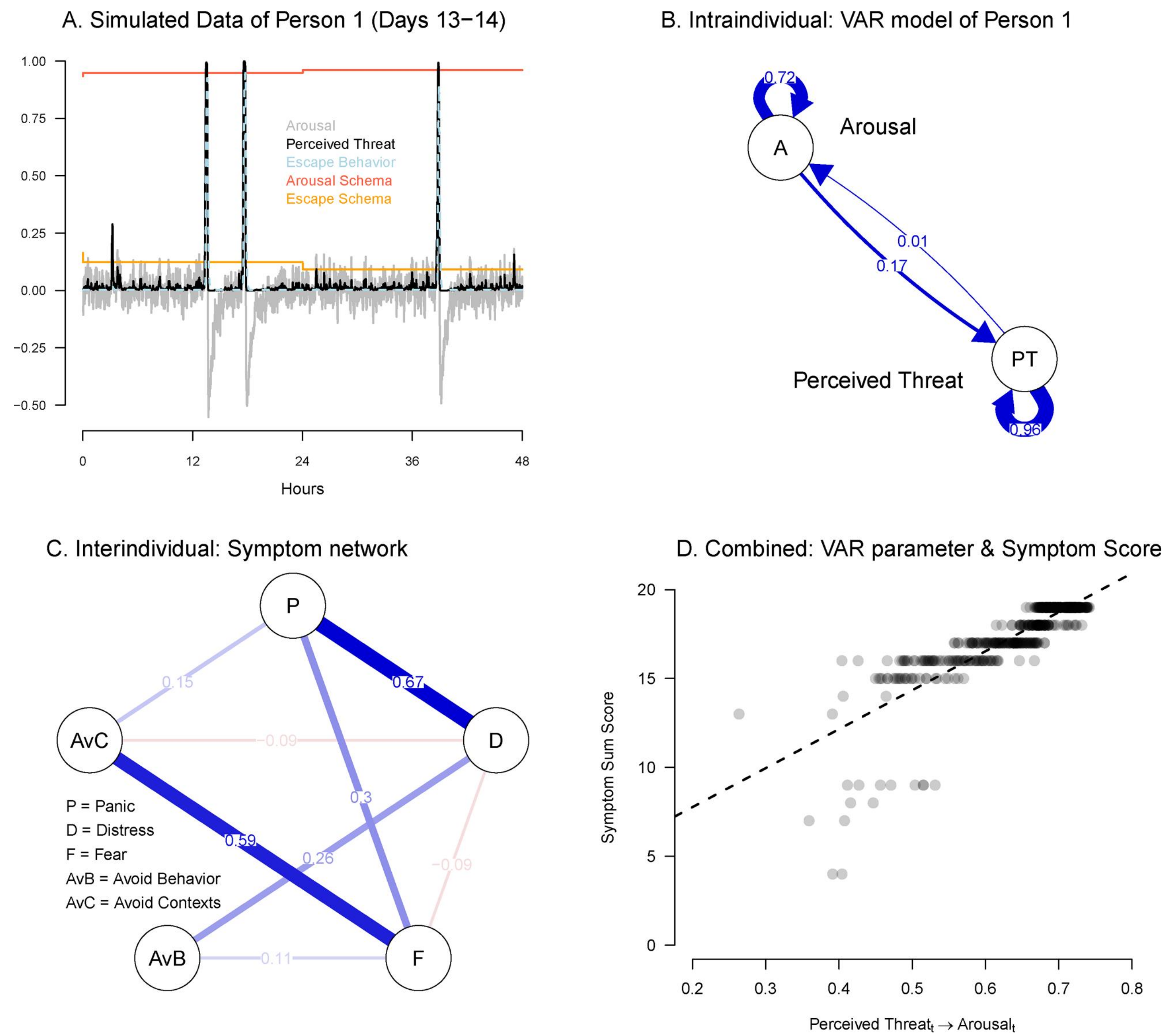

Figure 3. How the Panic Disorder model implies different types of phenomena. (a) Simulated data from days 14 and 15 of Person 1; we display the variables Arousal, Perceived Threat, Escape Behavior, Arousal Schema and Escape Schema; (b) A VAR model fitted to the Arousal (A) and Perceived Threat (PT) values measured at a minute time scale from Person 1; (c) Partial correlations between five symptoms, estimated based on a cross-sectional dataset consisting of the 500 simulated individuals; (d) The bivariate distribution of the lagged effect of PT on A and the symptom sum score in the last week in the sample of the 500 simulated individuals.

four weeks of data from $N = 500$ persons. The persons differ in their initial values of Arousal Schema and Escape Schema, which are chosen such that about 20% of individuals develop panic disorder within the four weeks. The data of days 13 and 14 of Person 1 are displayed in Panel A of Figure 3. We see that Arousal and Perceived Threat display variation at low levels until noon on day 13. This variation represents natural variation in Arousal, for example due to walking up the stairs or drinking coffee. However, at around 13h a perturbation of Arousal is large enough to kick off an escalating feedback loop between Arousal and Perceived Threat which culminates into a panic attack. After 5-20 min, the homeostatic feedback mechanism kicks in and brings Arousal (and therefore also Perceived Threat) back to normal levels. Since Escape Schema is low, Person 1 immediately shows Escape Behavior once Perceived Threat is elevated. Once Perceived Threat is low again, Escape Behavior also goes back to normal levels. We see two more panic attacks, one on the evening of day 13 and another on the evening of day 14.

The variables Arousal Schema and Escape Schema change at a slower time scale than Arousal, Perceived

Threat and Escape Behavior. We see that they are constant during day 13. However, Person 1 experienced two panic attacks on that day and showed Escape Behavior. They therefore learned that Arousal has potentially dangerous consequences, which were only avoided by showing Escape Behavior. As a consequence, Arousal Schema increases and Escape Schema decreases after day 13. On day 14 another panic attack occurs during which Person 1 shows Escape Behavior, which means that Arousal (Escape) Schema will again increase (decrease) during the next day. This shows that Person 1 is on the path of developing increasingly severe symptomatology that moves toward Panic Disorder. In fact, we initialized Person 1 with an Arousal Schema value of 0.812 and an Escape Schema value of 0.119, which means that a number of panic attacks must have occurred already before day 13. The code to repeat the simulation and reproduce the results shown in Figure 3 can be found at https://github.com/jmbh/withinbetweentheory.

Deriving intraindividual, interindividual, and combined phenomena

One way to examine the predictions of the computational model is to visualize the generated time series data and discuss its behavior qualitatively, as we have just done. In this way, we could also compare the simulated time series with a corresponding empirical time series, and thereby evaluate to what extent the computational model is faithful to reality. While eyeballing the data is often useful, it has the obvious limitations that it is not transparent or reproducible, and that it provides no principled way to separate signal from noise. We therefore typically summarize time series data in statistical models, which we see as tools to establish empirical phenomena (Borsboom et al., 2021; Haslbeck et al., 2022).

Intraindividual phenomenon

Focusing again on Person 1, we can summarize the time series by assessing how strongly variables are correlated with each other across a certain timespan. For example, we might be interested in extent to which Arousal at time t is correlated with itself one minute later at time point $t+1$. Extending such relationships to several variables leads to the Vector Autoregressive (VAR) model (Hamilton, 1994), which jointly models all variables at t as a linear function of all variables at one or several time points before, e.g., $t-1, t-2, \ldots$. In panel B of Figure 3 we display the conditional auto- and cross-correlations of a lag-1 VAR model estimated from minute-level measurements of Arousal and Perceived Threat of Person 1. Since these relationships are created by contrasting (or relating) measurements of the same subject, we consider this VAR model to represent an intraindividual phenomenon.

How would we use this intraindividual phenomenon to develop a theory of panic disorder? First, we need to estimate the corresponding VAR model from empirical data. These empirical data are sampled from Person 1 in the case of truly idiographic research, or a population of individuals that can considered interchangeably. Since we know the data generating model, we know that the VAR model does not correctly specify the data generating model in its full complexity. There are several reasons for this: Variables are omitted (e.g., Arousal Schema), the discrete minute-time scale does not match the continuous time of the generating model, and the VAR model is functionally misspecified in that it only includes linear relationships, while the true relationships are non-linear. This means that we generally have no guarantees for making accurate *direct* inferences from the VAR parameters about the mechanics of the generating model (see also Haslbeck et al., 2022; Haslbeck & Ryan, 2022).

However, we can think of the VAR estimates as phenomena to be explained by a computational model of panic disorder, which in this way put *constraints* on such a model. For example, we could collect experience sampling data from a given individual and estimate a VAR model on that data. Similarly, we can simulate time series data from the current iteration of the computational model and estimate a model-implied VAR model on those simulated data. Now, if the model-implied VAR model in Panel B is very different from its empirical counterpart, we know that our computational model needs to be improved, and we can use our understanding of the model and the substantive matter to propose changes to the model that would imply VAR models that better match its empirical counterpart. Importantly, however, for this comparison to work, the measurement frequency in the empirical data (perhaps every 2h) and in the simulated data (in principle infinite for continuous-time models) must be matched. In this example, this can be achieved by subsampling the simulated data to match the empirical measurement frequency. See also Ryan et al. (2023a) who validate a computational model of emotion dynamics by evaluating whether it reproduces phenomena captured by VAR models found typically in empirical data.

This example illustrates an interesting point: even if the levels of theory and phenomenon are aligned such as in this case, we may need to resort to the constraint-based type of theory development. We think that this might be the norm rather than the exception in theory development due to the fact that most statistical models are gravely misspecified in most situations due to missing variables, functional misspecification or limitations of measurements.

Interindividual phenomenon

To derive an interindividual phenomenon, we consider data from all of the 500 simulated persons. For each person, we determine the five symptoms experience of panic attacks (P), distress during panic attacks (D), fears related to panic attacks (F), avoidance of behaviors (AvB), and avoidance of contexts (AvC) from the Panic Disorder Severity Scale (PDSS; Houck et al., 2002). The symptoms are assessed in each of the four weeks, and are scored on a 0-4 Likert Scale. The symptoms are scored by creating a mapping from the minute-level data stream of all variables within a given week to the five answer categories of the ordinal scale. We did this for each of the five symptoms in a way that we found is most faithful to the precise wording of the symptom in the PDSS. The exact mapping can be found in Ryan et al. (2023b) and our reproducibility archive (https://github.com/jmbh/within-inbetweentheory). We then estimate a Gaussian Graphical Model (GGM) on the symptom scores in the last week using the graphical lasso (Epskamp et al., 2018; Friedman et al., 2008). Panel C in Figure 3 displays the partial correlations of the GGM. Because the variables in the model are defined by contrasting individuals to each other, we consider the symptom network in panel C an interindividual phenomenon. Note that the variables defined in this way integrate symptomatology over a time period (a week, in this case), which means that time information is present in the scores, even if these are contrasted interindividually. We will return to this implicit use of time information in the discussion.

Similarly to the intraindividual phenomenon, this interindividual phenomenon can be used as a constraint on theories about panic disorder that include inter-individual differences. For instance, if we estimated a GGM on corresponding empirical symptom data and found a much stronger relationship between the experience of panic attacks (P) and the avoidance of contexts (AvC), we would need to adapt the computational model to account for such a strong relationship. Importantly, this change might be made in a part of the model specifying the intraindividual dynamics of every individual in the model, which in interaction with inter-individual differences makes predictions about the interindividual correlation between the symptoms P and AvC.

A concrete example of how interindividual phenomena can guide the development of an intraindividual theory can be found in Haslbeck et al. (2022), where the authors simulated from an earlier version of the panic model of Robinaugh et al. (2019) which showed that panic attacks *always* co-occurred with the panic disorder symptoms persistent worry and avoidance behavior. However, when analyzing a large epidemiological survey (Alegria et al., 2007), they found that some people have occasional panic attacks but do not develop any further symptoms of panic disorder. This mismatch between intraindividual theory and interindividual phenomenon led to the current version of the model which includes the additional component Escape Schema (see Figure 2).

Combined phenomenon

To derive a combined phenomenon implied from the model, we again consider the data of all $N = 500$ subjects. We estimate the VAR model shown in panel B for each subject, and also record the PDSS symptom sum score of each subject in the last week. Panel D of Figure 3 displays the distribution over the lagged effect of Perceived Threat at time $t-1$ on Arousal at t and the PDSS score. We see that individuals with a stronger statistical association between Perceived Threat at $t-1$ on Arousal at t tend to have higher symptom sum score in the last week. This is roughly consistent with the general dynamics of the data generating model, in which a stronger feedback loop between Arousal and Perceived Threat leads to higher vulnerability to Panic Attacks. To define this phenomenon *via* the probability distribution over the VAR parameter and the symptom sum score we first used many intraindividual contrasts to obtain the VAR coefficients; and then used interindividual contrasts to create the bivariate distribution in panel D. We therefore refer to this phenomenon as a combined phenomenon. Similar to the intra- and interindividual-phenomena discussed above, this phenomenon can be used in theory development in the sense of a constraint: If the corresponding bivariate distribution in empirical data looks very similar, we would take this as evidence for the adequacy of our computational model; if not, we would use it to propose changes that lead to mode-implied distributions that are closer to the empirical one.

Discussion

The question of how our understanding of individual persons relates to different types of data that use intra- and interindividual contrasts is central to psychological science. Molenaar (2004) put the issue squarely on the methodological research agenda, and his paper has been an important motivation for the development of techniques suited to address intraindividual data on their own terms (Bringmann et al., 2017; Epskamp et al., 2018; Fisher et al., 2017; Gates & Molenaar, 2012; Hamaker, 2012). In combination with the ever increasing technological possibilities for gathering data through digital devices (Conner & Barrett, 2012; Miller, 2012; Trull & Ebner-Priemer, 2014), this movement is reshaping the field of scientific psychology rapidly.

However, even though the relation between statistical models and data is quite clear in its implications for how the intra- and interindividual levels relate (Molenaar, 2004; Molenaar et al., 2003), the relation between scientific theories and data is less straightforward. In particular, it is not always clear how different types of data can inform theories, and a systematic framework of studying this question is lacking. In the present paper, we have proposed that such a framework can be realized through the combination of the theory-phenomena-data distinction and the formalization of theories. Formalized theories can be used to derive implications in the form of empirical phenomena—robust statistical patterns—that we should expect to see if the theory were true. This feature makes it possible to answer the question of whether and how different kinds of data may weigh in on a given theory, by considering how the phenomena they evidence constrain the space of candidate theories.

Our analysis of the computational model of panic disorder demonstrates that a psychological theory can imply both intra- and interindividual phenomena. If a theory is able to do this, it can draw evidence from either of these sources in the form of giving successful explanations of intraindividual, interindividual, or combined phenomena. Conversely, each type of phenomenon can constrain the theory, in the sense that it limits the degree to which the theory can be varied without losing the explanatory connection to the phenomena. This aligns with the hard-to-vary principle proposed by Deutsch (2012): the degree to which one can vary functional details of the explanation is inversely proportional to the explanatory power of the theory. Using this focus on explanation, researchers may simulate implications of theories in various data types. In this way, we may be able to determine, on a case by case basis, whether they would gain better information from gathering and analyzing time series, mean differences, experimental effects, or interindividual correlations.

Importantly, our characterization of intraindividual, interindividual, and combined phenomena is framed in terms of comparisons that a researcher makes. However, there are also comparisons that participants engage in when they answer psychometric items. For example, questions that require integration over information over time windows arise often in cross-sectional measurements (Ryan and Dablander, n.d.), for instance when psychopathology researchers ask a person how well they have slept over the past two weeks, or when personality researchers probe how often participants engage in certain behaviors in general. And even in time series measurements, questions are often not with respect to the current state but with respect to the time period since the last measurement (Haslbeck & Ryan, 2022). This suggests that a subset of the phenomena we label as interindividual phenomena may be *implicitly combined*, because the representational content of the data involves an implicit integration over a time domain. How to deal with this issue is an important open question in this domain, and while it is beyond the score of the present paper, we think this problem deserves more psychometric scrutiny than it currently enjoys.

Another interesting issue is the question of where causal effects should be situated. Often, research into causal effects uses interindividual contrasts (e.g., in a simple Randomized Controlled Trial with a single measurement occasion). Because, in the current framework, we have defined empirical phenomena as statistical patterns in data, this means that observed mean differences between experimental conditions classify as interindividual phenomena. The interpretation of these mean differences in terms of causal effects then rests on the explanation of interindividual phenomena in terms of an intraindividual theory. Thus, intraindividual causal theories explain interindividual phenomena (mean differences between groups), and as a result the interindividual phenomena form evidence for these theories. Now, as the evidence mounts, so that the experimental findings become highly robust and replicable (e.g., think of the speed-accuracy tradeoff or the Stroop effect in psychology), then the inferred intraindividual causal relations can themselves become explanatory targets of a more expansive theory. In this way, the existence of mechanisms that were originally posited in the context of an explanatory theory can become a phenomenon to be explained in the context of a deeper explanatory theory. Examples of such changes abound in science; for instance, although germs were originally posited as hypothetical entities that explained disease transmission, currently they are so well evidenced that deeper theories can be developed to explain their existence; as a result, the existence of germs has changed from a hypothesis in a tentative explanatory theory to a robust fact that can itself be a target of explanation. One can imagine an explanatory hierarchy of empirical phenomena in which the intuitive interpretation of phenomena changes, for instance from interindividual to intraindividual, when the evidence is considered so secure that the intraindividual explanation of interindividual phenomena attains the status of empirical fact, which may itself be explained by a deeper theory. In terms of such a hierarchy, the type of empirical phenomena we have worked with here would be situated at the relatively low level of empirical generalizations (Haig, 2008). Further research is needed to evaluate the structure of such a hierarchy and to investigate its consequences for thinking about the relation between intra- and interindividual theories and phenomena.

As we have shown, if one has a formalized theory, one can evaluate a) which intra- and interindividual phenomena the theory implies, and b) how such phenomena can inform intra- and interindividual parts of the theory. This has direct implications for the organization of research and the connection between the two disciplines Cronbach (1957) identified. For example, if the theory implies that two variables should be correlated intraindividually, an ESM design can be combined with an Vector Autoregressive model. If the theory implies that randomized interventions should produce mean differences, an experimental study can be used in combination with an analysis of variance. If the theory implies that individual differences on questionnaire items should be unidimensional, a correlational study can be combined with a factor analysis to study this. Importantly, as the panic disorder example shows, one and the same theory can carry implications in all of these directions. Therefore, such theories should be able to form structures that connect the methodological traditions of psychology, because they can serve as focal points for the interaction between theorists and empirical researchers and coordinate research efforts systematically.

Importantly, these bridges between traditions are afforded by the addition of phenomena as intermediaries between theories and data (Bogen & Woodward, 1988; Woodward, 2011). We think that a focus on this level may lead to improvements in the methodology of indirect inference (Haslbeck et al., 2022; Hosseinichimeh et al., 2016). In our view, such approaches should play an important role in future psychological research, as the discipline moves toward the construction of stronger psychological theory (Borsboom et al., 2021; Fried, 2020; Robinaugh et al., 2021; van Rooij & Baggio, 2021).

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: Denny Borsboom is supported by NWO Vici grant no. 181.029. JH has been supported by the gravitation project 'New Science of Mental Disorders' (www.nsmd.eu), supported by the Dutch Research Council and the Dutch Ministry of Education, Culture and Science (NWO gravitation grant number 024.004.016).

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Adolf, J., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, *5*, 883. https://doi.org/10.3389/fpsyg.2014.00883

Alegria, M., Jackson, J. S., Kessler, R. C., & Takeuchi, D. (2007). *Collaborative psychiatric epidemiology surveys (CPES), 2001-2003 [United States]*. Inter-university Consortium for Political and Social Research.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://

apastyle.apa.org/style-grammar-guidelines/references/examples/diagnostic-manual-references

Anderson, J. R. (1996). Act: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355–365. https://doi.org/10.1037/0003-066X.51.4.355

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022). Word-object learning via visual exploration in space (wolves): A neural process model of cross-situational word learning. *Psychological Review*, *129*(4), 640–695. https://doi.org/10.1037/rev0000313

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, *97*(3), 303–352. https://doi.org/10.2307/2185445

Borsboom, D., Haslbeck, J. M., & Robinaugh, D. J. (2022). Systems-based approaches to mental disorders are the only game in town. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *21*(3), 420–422. https://doi.org/10.1002/wps.21004

Borsboom, D., Kievit, R., Cervone, D. P., & Hood, S. B. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In J. Valsiner, P. Molenaar, M. Lyra, & N. Chaudary (Eds.), *Developmental process methodology in the social and developmental sciences*. Springer.

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J. P., van der Krieke, L., de Jonge, P., & Wichers, M. (2017). Can we jump from cross-sectional to dynamic interpretations of networks? Implications for the network perspective in psychiatry. *Psychotherapy and Psychosomatics*, *86*(3), 175–177. https://doi.org/10.1159/000453583

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*(3), 409–425. https://doi.org/10.1037/met0000085

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods*, *49*(3), 988–1005. https://doi.org/10.3758/s13428-016-0754-9

Cattell, J. (1890). V.- Mental tests and measurement. *Mind*, *os-XV*(59), 373–381. https://doi.org/10.1093/mind/os-XV.59.373

Cattell, R. B. (1988). The Data Box. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 69–130). Springer.

Clark, D. M. (1986). A cognitive approach to panic. *Behaviour Research and Therapy*, *24*(4), 461–470. https://doi.org/10.1016/0005-7967(86)90011-2

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*(4), 327–337. https://doi.org/10.1097/PSY.0b013e3182546f18

Contreras, A., Nieto, I., Valiente, C., Espinosa, R., & Vazquez, C. (2019). The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, *88*(2), 71–83. https://doi.org/10.1159/000497425

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C. S., Sheeber, L., Allen, N., & Hamaker, E. L. (2017). On the use of mixed markov models for intensive longitudinal data. *Multivariate Behavioral Research*, *52*(6), 747–767. https://doi.org/10.1080/00273171.2017.1370364

Deutsch, D. (2012). *The beginning of infinity: Explanations that transform the world*. Penguin Books.

Devitt, M. (2005). Scientific realism. In F. Jackson & M. Smith (Eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford University Press.

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*(11), 988–994.

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*(8), 1044–1056. https://doi.org/10.1037/abn0000311

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Galton, F. (1879). Psychometric experiments. *Brain*, *2*(2), 149–162. https://doi.org/10.1093/brain/2.2.149

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*(1), 116–136. https://doi.org/10.1037/0033-295x.109.1.116

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*(1), 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Haig, B. D. (2008). Scientific method, abduction, and clinical reasoning. *Journal of Clinical Psychology*, *64*(9), 1013–1018. https://doi.org/10.1002/jclp.20505

Hamaker, E. L. (2012). Why researchers should think" within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2016). Modeling bas dysregulation in bipolar disorder: Illustrating the potential of time series analysis. *Assessment*, *23*(4), 436–446. https://doi.org/10.1177/1073191116632339

Hamaker, E. L., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated trait-state model. *Journal of Research in Personality*, *41*(2), 295–315. https://doi.org/10.1016/j.jrp.2006.04.003

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Hamilton, J. D. (1994). *Time series analysis* (vol. 2). Princeton University Press.

Haslbeck, J. M., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research*, *56*(1), 120–149. https://doi.org/10.1080/00273171.2020.1743630

Haslbeck, J. M. B., & Ryan, O. (2022). Recovering Within Person Dynamics from Psychological Time Series. *Multivariate behavioral research*, *57*(5), 735–766. https://doi.org/10.1080/00273171.2021.1896353

Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930–957. https://doi.org/10.1037/met0000303

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150. https://doi.org/10.3389/fnins.2014.00150

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Sage.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.2307/2289064

Hosseinichimeh, N., Rahmandad, H., Jalali, M. S., & Wittenborn, A. K. (2016). Estimating the parameters of system dynamics models using indirect inference. *System Dynamics Review*, *32*(2), 156–180. https://doi.org/10.1002/sdr.1558

Houck, P. R., Spiegel, D. A., Shear, M. K., & Rucci, P. (2002). Reliability of the self-report version of the panic disorder severity scale. *Depression and Anxiety*, *15*(4), 183–185. https://doi.org/10.1002/da.10049

Isvoranu, A. M., Van Borkulo, C. D., Boyette, L. L., Wigman, J. T., Vinkers, C. H., Borsboom, D., Kahn, R., De Haan, L., Van Os, J., Wiersma, D., Bruggeman, R., Cahn, W., Meijer, C., & Myin-Germeys, I. (2017). A network approach to psychosis: Pathways between childhood trauma and psychotic symptoms. *Schizophrenia Bulletin*, *43*(1), 187–196. https://doi.org/10.1093/schbul/sbw055

James, W. (1909). *The meaning of truth*. Dover Publications.

Jensen, A. R. (1999). *The g factor: The science of mental ability*. Praeger.

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., Tyler, L. K., Brayne, C., Bullmore, E., Calder, A., Cusack, R., Dalgleish, T., Matthews, F., Marslen-Wilson, W., Rowe, J., Shafto, M., Campbell, K., Cheung, T., Geerligs, L., McCarrey, A.; Cam-CAN Research team (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature Communications*, *5*(1), 5658. https://doi.org/10.1038/ncomms6658

Kievit, R. A., Frankenhuis, W. E., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, 513. https://doi.org/10.3389/fpsyg.2013.00513

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. I). Academic Press.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984–991. https://doi.org/10.1177/0956797610372634

Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some recommendations on the use of daily life methods in affective science. *Affective Science*, *3*(2), 505–515. https://doi.org/10.1007/s42761-022-00101-0

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

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *261*(5560), 459–467. https://doi.org/10.1038/261459a0

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

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 *Understanding human development: Dialogues with lifespan psychology* (pp. 339–360). Kluwer Academic Publishers, Molenaar, P. C. M.: Faculty of Psychology, University of Amsterdam.

Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the Random Intercept Cross-Lagged Panel Model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 638–648. https://doi.org/10.1080/10705511.2020.1784738

Murray, D. (2020). *The creation of scientific psychology*. Routledge.

Pearl, J. (2009). *Causality*. Cambridge University Press.
Peirce, C. S. (1931). *The Collected Papers of Charles Sanders Peirce* (vol. I–VI). Harvard University Press.
Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. The MIT Press.
Posthuma, D., De Geus, E. J., Baaré, W. F., 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*(2), 83–84. https://doi.org/10.1038/nn0202-83
Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420
Reynolds, C. W. (1987). *Flocks, herds and schools: A distributed behavioral model* [Paper presentation]. Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (pp. 25–34). https://doi.org/10.1145/37401.37406
Robinaugh, D. J., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., van Nes, E. H., Scheffer, M., Kendler, K. S., et al. (2019). Advancing the network theory of mental disorders: A computational model of panic disorder. https://doi.org/10.31234/osf.io/km37w
Robinaugh, D. J., Haslbeck, J. M. B., Waldorp, L. J., Kossakowski, J. J., Fried, E. I., Millner, A. J., McNally, R. J., Ryan, O., de Ron, J., van der Maas, H. L. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Borsboom, D. (2019). Advancing the Network Theory of Mental Disorders: A Computational Model of Panic Disorder. https://doi.org/10.31234/osf.io/km37w
Robinaugh, D. J., Hoekstra, R. H., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008–2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404
Rohrer, J. M., & Murayama, K. (2023). These are not the effects you are looking for: Causality and the within-/between-persons distinction in longitudinal data analysis. *Advances in Methods and Practices in Psychological Science*, *6*(1), 251524592211408. https://doi.org/10.1177/25152459221140842
Ryan, O., & Dablander, F. (2022). Equilibrium Causal Models: Connecting Dynamical Systems Modeling and Cross Sectional Data Analysis. https://doi.org/10.31234/osf.io/q4d9g
Ryan, O., Haslbeck, J. M. B., & Robinaugh, D. J. (2023). Improving treatments for mental disorders using computational models. https://doi.org/10.31234/osf.io/tyn9z
Ryle, G. (1949). *The concept of mind*. Hutchinson.
Smaldino, P. E. (2017). Models are stupid, and we need more of them. In *Computational social psychology* (pp. 311–331). Routledge.
Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–293. https://doi.org/10.2307/1412107
Suppe, F. (1977). *The structure of scientific theories*. University of Illinois Press.
Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis.
Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, *23*(6), 466–470. https://doi.org/10.1177/0963721414550706
Tuerlinckx, F., & De Boeck, P. (2005). Two interpretations of the discrimination parameter.
van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 87–92. https://doi.org/10.1073/pnas.1312114110
Van der Maas, H., & Molenaar, P. C. M. (1992). A catastrophe-theoretical approach to cognitive development. *Psychological Review*, *99*(3), 395–417. https://doi.org/10.1037/0033-295x.99.3.395
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*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842
van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, *118*(2), 339–356. https://doi.org/10.1037/a0022749
van Dongen, N., van Bork, R., Finnemann, A., van der Maas, H., Robinaugh, D., Haslbeck, J., de Ron, J., & Borsboom, D. (2022). Improving psychological explanations.
van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(4), 682–697. https://doi.org/10.1177/1745691620970604
Wagenmakers, E.-J., van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. https://doi.org/10.3758/bf03194023
Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician*, *36*(1), 46–48. https://doi.org/10.1080/00031305.1982.10482778
Weinberger, N. (2015). If intelligence is a cause, it is a within-subjects cause. *Theory & Psychology*, *25*(3), 346–361. https://doi.org/10.1177/0959354315569832
Wichers, M., & Groot, P. C..; Psychosystems, ESM Group, and EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, *85*(2), 114–116. https://doi.org/10.1159/000441458
Wichers, M., Smit, A. C., & Snippe, E. (2020). Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: A confirmatory

single-subject time-series study. *Journal for Person-Oriented Research*, *6*(1), 1–15. https://doi.org/10.17505/jpor.2020.22042

Wichers, M., Wigman, J., & Myin-Germeys, I. (2015). Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review*, *7*(4), 362–367. https://doi.org/10.1177/1754073915590623

Wijsen, L. D., Borsboom, D., Cabaço, T., & Heiser, W. J. (2019). An academic genealogy of Psychometric Society presidents. *Psychometrika*, *84*(2), 562–588. https://doi.org/10.1007/s11336-018-09651-4

Woodward, J. F. (2011). Data and phenomena: A restatement and defense. *Synthese*, *182*(1), 165–179. https://doi.org/10.1007/s11229-009-9618-5